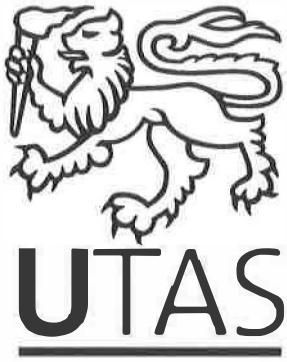
Feature Selection for Prosthetic Control using Myoelectric Signals

by

Le Minh Diep Khong, B.E. Hons (UTAS)

Submitted in fulfilment of the requirements for the degree of
Master of Engineering Science by Research

School of Engineering
University of Tasmania
April, 2016



I declare that this thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signed: _____

Le Minh Diep Khong

Date: 13/04/2016

Signed: _____

Le Minh Diep Khong

Date: 13/04/2016

ABSTRACT

This thesis investigates selection of time domain (TD) signal features for myoelectric signal (MES) based control of motorised hand and wrist prostheses. A signal feature represents a distinguishing property of a MES to be used in pattern recognition algorithms. In particular, TD features reflect the mathematical functions and physical expression of the transient signal waveform with respect to time. Extracted features capture the structural details of a MES, minimise loss of information upon conversion, and simplify movement classification. The advantage of TD features is that they produce lower dimensional input vectors while maintaining sufficient accuracy of various movements if adequate information is provided. Feature sets as a solution to gather information in MES based control has not been thoroughly studied in the literature. We aim to develop methods to elevate the use of TD features and suggest a comprehensive feature set that is helpful in pattern recognition.

Myolectric signals used in this study were from the BioPatRec database, an open source platform for research and control of artificial limbs via pattern recognition using bioelectric signals. This database is named as *10mov4chUntargetedForearm* comprising data on10 hand and wrist movements acquired by 4 bipolar sEMG channels from the left or right forearm.

Based on feature selection (FS) which preserves information of the MES, we propose three methods, namely a genetic algorithm (GA), class relevant criteria and a self-organising feature map (SOFM) to assemble feature sets from TD features of twenty one candidates. To evaluate these feature sets, we implemented three pattern recognition algorithms, particularly the Multilayer Perceptron (MLP), Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) algorithms. The reported movement accuracy and Wilcoxon p value demonstrated that the proposed feature sets consistently outperformed a typical feature set found in the literature; in particular, improved the accuracy of poor quality datasets from 85% to 93%.

The thesis has made a thorough investigation of TD features contributing in three categories. Firstly, we developed a variety of independent methods for FS. It was noticed that FS has been limited in meta-heuristic searches in the literature. We have demonstrated that there are several solutions that use potential TD features to assemble a feature set to be used in pattern recognition. Secondly, we have shown that statistical tests can be successfully applied in FS. Thirdly, we explored an investigation of data along time series vectors instead of analysing it conventionally by time segmentation. The success of this method suggests a new way that may value further inspection.

In brief, this thesis presents possible solutions for TD feature based pre-processing of the input of pattern recognition algorithms for prosthetic control. It provides

immediate accuracy improvement through a replacement of feature sets and further implementation in methodology for FS.

ACKNOWLEDGEMENTS

This thesis has been achievable with great supports and encouragements from many people. Firstly, I would like to express my gratitude to my primary supervisor, Dr Timothy Gale, for his patience and confidence in my abilities. He is a great mentor whose examines and supervises me through hardships I encountered in my degree. I also appreciate my co-supervisors, Prof. Jan C. Olivier and Dr Danchi Jiang for their encouragements and continuous supports.

I would like to thanks to Max Ortiz-Catalan who inspired me to work on this project. This thesis would not be completed without data sources as well as supplementary software provided by Max and his colleagues.

I would like to thanks to all my colleagues and friends whom I have shared knowledge and experience during the time of my study.

Lastly, it could not be possible without the unconditional love and eternal encouragements of my parents and family. This thesis is dedicated to all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AG	thumb up
CH	hand close
CP	contribution percentage
CS	class separability
DB	Davis-Bouldin index
EH	wrist extension
FD	frequency domain
FG	fine grip
FH	wrist flexion
FP	feature projection
FS	feature selection
GA	Genetic Algorithm
IDM	input difference from mean
LDA	Linear Discriminant Analysis
MES	myoelectric signal
MI	mutation information
MLP	Multilayer Perceptron
MRMR	minimal-redundancy-maximal-relevance
OH	hand open
PCA	principle component analysis
PN	hand pronation
PT	pointing
sEMG	surface electromyography

SG side grip

SN hand supination

SOFM Self-Organising Feature Map

SVM Support Vector Machine

t test Two-sample t-test

TAC Target Achievement Control

TD time domain

TFD time-frequency domain

TSCD time-scale domain

TSD time-serial domain

U test Mann-Whitney U test

Chapter 1

Introduction

1.1 Introduction

This thesis introduces alternative approaches of generating myoelectric signal (MES) feature sets for surface electromyography (sEMG) based control of motorised prosthetic hands and wrists. Typical feature sets currently used have been shown to achieve high accuracy in hand motion pattern recognition [34]. However, there is still potential to improve accuracy by using comprehensive feature sets that have not already been investigated in the literature. The focus of this thesis is on finding alternative time domain (TD) feature sets that are of low dimension and maintain or improve accuracy.

This chapter introduces the overall research area, including prostheses, pattern recognition algorithms and types of signal features in the literature, as well as the specific signal features used in the research presented in this thesis. This chapter also references relevant introductory literature, with a review related to dimensionality reduction given later in chapter 2 and reference to other relevant literature given in chapters 3, 4 and 5.

1.2 Prosthetic control

This research study considers myoelectric control of motorised prostheses. These prostheses have several advantages in terms of moderation, functionality, strong grip and simplicity of operation. The myoelectric signals are collected from the peripheral nervous system, illustrated in Fig. 1.1.

We use sEMG signals that have been recorded and made available in the BioPatRec [59] database. BioPatRec is an open source research platform built to enable research of algorithms for prosthetic control, including algorithms for sEMG signal pattern recognition. Data is shared under a *Data Respository* including the dataset a *10mov4chUntargetedForearm* used in this thesis which can be downloaded from BioPatRec webpage [58]. This dataset contains data of 10 hand and wrist motions acquired by 4 bipolar electrodes attached on the skin of the untargeted forearm (i.e.

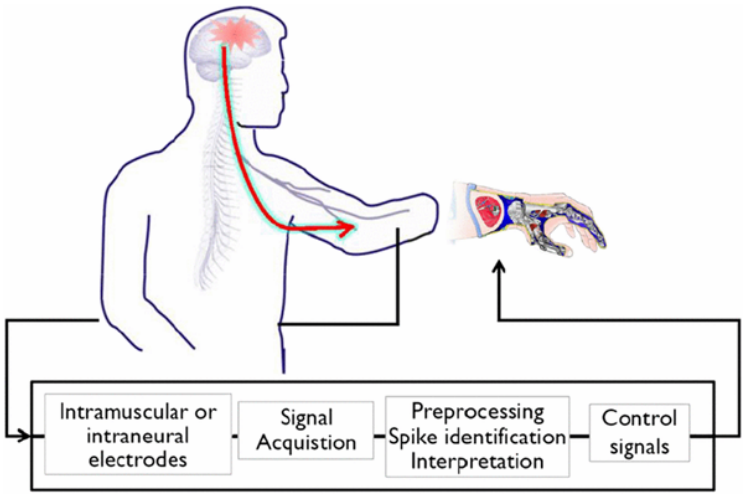


Figure 1.1: Schematic representation of hand prosthesis control paradigm [53].

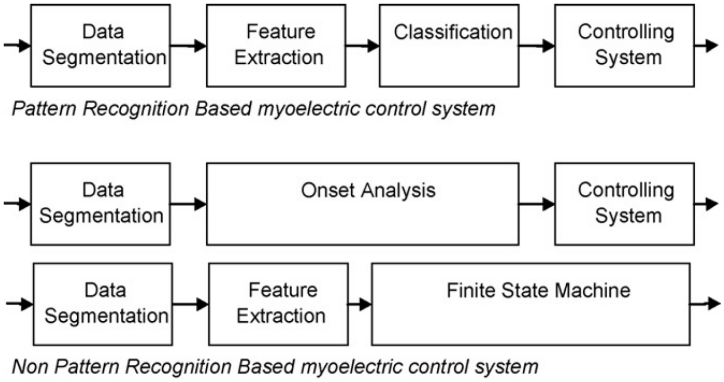


Figure 1.2: Two main approaches for prosthetic control based on myoelectric signals [4].

either left or right forearm), collected from 20 voluntary subjects of good health and no amputation. In summary, each subject watched a visual arm illustrating 10 movements and was then instructed to repeat the movements consecutively with 3 repetitions of each movement.

Myoelectric control can be categorised into pattern recognition and non-pattern recognition (Fig. 1.2). The former approach discriminates patterns using classifiers in which the performance is highly dependent on the algorithms. The latter approach separates the desired classes by a sequence of input signals. Due to the simple structure and the dependence on the characteristic of sequential signals its functionality is limited. The pattern recognition approach, on the other hand, is flexible in choosing an approximate class discrimination method. The algorithm can be optimised by tuning parameters which are independent of the input source. Hence, we evaluate the performance of feature sets by using the pattern recognition approach.

Current methods to measure prosthetic control performance in real-time includes

the Motion Test [44] and Target Achievement Control (TAC) [68]. The Motion test measures the recognition time of correct movements and prediction accuracy. The TAC requires a virtual reality environment to simulate a prosthetic device so as to measure a target position and misclassification (i.e. overshoot the target posture due to unintended movement). They are included in BioPatRec [59].

Because this thesis uses pre-recorded data provided from BioPatRec, this work is completed offline and measures the performance of pattern recognition by the classification accuracy. The offline accuracy is

$$acc = \frac{\text{Number of correctly classified samples}}{\text{Total number of testing samples}} \times 100\%.$$

1.3 Pattern recognition algorithms

Pattern recognition algorithms typically assign signal patterns to pre-determined movement classes in order to convert the correct predictions to electrical command signals to the devices [59, 34]. Technically, the output of the algorithms is a binary string representing a given number of classes in which the assigned class is given the value 1 and the rest given the value 0. Selection of an appropriate algorithm depends on its ability to accommodate feature variation [34]. This happens when collecting signals from different patients, due to a range of factors such as congenital defects, the position of electrodes and fluctuations in body weight [1, 12]. The other requirements to select an algorithm is the ability to quickly respond to a command with a high level of accuracy, with a response time typically of less than 300 ms [44, 46]. Conventional pattern recognition algorithms which satisfy these conditions include the Multi-layer Perceptron (MLP) [34], Linear Discriminant Analysis (LDA) [46] and Support Vector Machine (SVM) [60]. In this thesis, these classifiers are used to evaluate the performance of several sets of myoelectric signal features.

1.4 Signal features and feature sets

1.4.1 Signal features

Myoelectric signal features are used to generate feature vectors which are the input of pattern recognition algorithms. The contemporary approach to extract these features is signal segmentation, in which a series of signals are cut into several time segments [34]. Features are computed by measuring specific factors of the raw signals such as RMS values or frequency and averaging them within a specific time segment. Several types of features have been studied [78, 4, 60, 77, 24] which can be categorised into four groups: time domain (TD), time-serial domain (TSD), frequency or spectral domain (FD), and time-scale or time-frequency domain (TSCD or TFD).

The most popular type of features are TD which were first introduced in 1993 by Hudgins [34]. This type typically reflects the amplitude of signals which can be easily extracted from each time segment with a simple mathematical computation.

Therefore, they are widely used in real-time control due to the short extraction time required. Commonly used features include the mean absolute value (*tmabs*) [34, 77, 63], integrated absolute value (*tiav*) [54], variance (*tvar*) [63, 70], mean absolute value slope (*tmavs*) [34], Willison amplitude (*twamp*) [70], zero crossing (*tzc*) [34], slope sign changes (*tssc*) [26], waveform length (*twl*) [26], and EMG histogram (*tehist*) [77].

TSD features examine the stability of signals in a time series. Typical TSD features are autoregressive coefficient (*tsar*) [54, 27, 28, 35], and Cepstral coefficient (*tscc*) [63].

FD features measure the intensity of muscle contraction from the analysis of the MES with respect to frequency. Typical FD features are power spectrum (*fps*) [60], mean and median frequencies (*fmn, fmd*) [60], frequency ratio (*fr*) [78].

TFD features specialise in bridging time information and spectral information which is useful when the signals are multiple time-varying frequencies. Typical TFD features are short-time Fourier transform (*tfstft*), wavelet transform (*tfwt*) and wavelet packet transform (*tfwpt*) [23, 24].

1.4.2 Feature sets

A feature set is a group of multiple features used for a multi-function control purpose. It extracts feature vectors, which are the input for the pattern recognition algorithms.

There are several factors which determine an optimal feature set. The most important factors are the class discrimination and the computational complexity [34]. The class discrimination measures the number of movements a feature set can correctly classify while the computational complexity measures the required time to extract that set. Several studies [34, 60, 59] show that TD features are able to satisfy these requirements.

This study proposes FS approaches based on TD features. We consider twenty one candidates that are commonly recommended in the literature and currently released in BioPatRec [59]:

1. Mean value (*tmn*)
2. Mean absolute value (*tmabs*)
3. Median (*tmd*)
4. Standard deviation (*tstd*)
5. Variance (*tvar*)
6. Waveform length (*twl*)
7. Root mean square (*trms*)
8. Zero crossing (*tzc*)

9. The number of peak values that overs RMS (*tpks*)
10. Mean of peaks over RMS (*tmpks*)
11. Mean of difference of peaks over RMS or the velocity of peaks (*tmvel*)
12. Slope sign change using tmvel value as threshold (*tslpch1*)
13. Slope sign change using tmabs value as threshold (*tslpch2*)
14. Power of waveform (*tpwr*)
15. Correlation of data between channels (*tcr*)
16. Covariance of data between channels (*tcv*)
17. Mean of the absolute difference (*tdam*)
18. Fractal dimension using Katz's algorithm (*tfd*)
19. Maximum fractal length (*tmfl*)
20. Fractal dimansion according to Higuchi (*tfdh*)
21. Rough entropy per channel (*tren*).

We propose three approaches to assemble small TD feature sets including approximately three to four features with minimal loss of accuracy. The performance of these sets is compared to that of a typical set found in the literature. This feature set demonstrated in the literature outperformed other feature sets in both time and frequency domains as well as adapted well to various lengths of time segment [60, 34, 59, 53].

1.5 Contributions

This thesis makes an investigation of FS with respect to TD features of sEMG signals intended for prosthetic hand control. The contributions can be summarised as follows:

- Development of three independent methods for FS which do not require repetition of dimensionality reduction algorithms. Our methods gather superior signal features used as optimal feature sets for the input of pattern recognition algorithms.
- The first application of statistical analysis in FS to search for superior signal features. This method shows the advantages in achieving high accuracy of pattern recognition.
- An introductory investigation of TD features along a time series. Conventionally, FS methods examine signal features by time segments. Studying TD features in a transient waveform allows us to analyse the response of signal features with respect to the variation of the MES during the entire time period of muscle activity.

1.6 Thesis synopsis

Below is a brief summary of each of the 6 chapters.

Chapter 1 provides an introduction to the general research area of pattern recognition based prosthetic control and the challenges associated with finding an optimal set of myoelectric signal features.

Chapter 2 reviews the literature which focuses on dimensionality reduction. Two major areas are considered: feature selection (FS) and feature projection (FP). The FS is suitable for TD features due to the ability to preserve signal information while the FP is applicable to other types of signal features due to the ability to reduce high dimensional input vectors produced from these features. Since our particular concern is TD features, FS is studied to generate optimal feature sets.

Chapter 3 presents a method for selecting optimal sets of features for each individual dataset based on the Genetic Algorithm (GA). GA searches along twenty one feature candidates and randomly selects a few of them for testing. Testing is done by using a Multi-layer Perceptron (MLP) which outputs the accuracy of hand and wrist movements. The set corresponding to the highest accuracy is preserved for reproduction and the cycle repeats until the terminal condition reached.

Chapter 4 introduces class relevant coefficient criteria including Max-Relevance and Min-Redundancy used in mutual information (MI); the other criteria are the Two-sample t-test (t test), and Mann-Whitney U test (U test). These criteria are used separately, pooling different sets of features. MI investigates the relationship between signal features, also between a signal feature and a target class. The t test calculates a dependent index of a signal feature with a target class, and assumes that the MES are normally distributed. The U test works in the same manner as the t test but assumes the MES are non-normally distributed.

Chapter 5 proposes finding the optimal FS using a Self-Organising Feature Map (SOFM) which investigates the MES in transient waveforms. Firstly, similar features are mapped to a cluster where a superior feature will be determined by feature evaluation. Secondly, the best signal feature candidates from many clusters are pooled into a set, and feature evaluation is replicated to calculate a contribution percentage (CP) of individual signal features. Subsequently, they are ranked according to their CPs. Finally, the top highest ranked signal features are selected so as to generate a feature subset that maintains 95% accuracy of the full set.

Chapter 6 concludes the thesis including the significant achievements, major contributions, and future work.

Chapter 2

Literature review and project objectives

2.1 Introduction

This chapter provides the background of the thesis. Firstly, we review relevant literature on the methodology of producing feature vectors. Secondly, methods for implementation of the input stage for pattern recognition algorithms are introduced, including approaches in dimensionality reduction. Subsequently, a summary of current methods outlined in the literature is given which suggests a solution for reduction of the input space. Finally, the research aims are stated, introducing three alternative methods for feature selection.

2.2 Feature vectors

The application of signal features in prosthetic control was first introduced by Hudgins et al. [34] in 1993. A feature vector is a combined vector of signal features in n-dimensional space. It is the input vector of pattern recognition algorithms. The vector is extracted from a single time window of transient signals. A feature vector carries information about a segment of signals reflecting a deterministic state and class of movement. In pattern recognition, the process of computation of signal features is termed feature extraction. Signal features are calculated in a particular form or mathematical function one-by-one in a segment of the MES. If signal features are averaged over the entire transition period, most of the structural details embedded in the MES will be lost [34]. Therefore, signal features are computed in multiple time segments using statistical approaches. This allows multiple outputs for sEMG prosthetic control from data originating from multiple channel sEMG.

The procedure of feature extraction can be summarized as follows. Firstly, the transient periods of signals are removed, and the entire or a part of muscle contraction waveform (typically 70% of the contraction time) are joined (Fig. 2.1b). Secondly, the signal is cut into many time segments (with or without overlap); each segment stores the same number of samples of data. Subsequently, many signal features are

computed from the same time segments and are arranged in chronological order. Finally, feature vectors are produced by combining extracted features. In other words, a feature vector represents a sample of time, in which several signal features are placed next to the others that converts the original data to more meaningful information.

2.3 Feature optimisation methods

The use of myoelectric signal features in prosthetic control has been widely applied in the literature because of its ability to allow multi-functional control using the same source of sEMG without increasing the effort made by amputees [34, 48]. It is also able to easily incorporate into new control schemes [53, 45, 11]. Because of an expansion of signal features, a demand for methodology to select appropriate signal features arises so as to limit feature extraction time as well as eliminate the increment of the input dimensionality [36].

The theoretical foundation of contemporary prosthetic control in signal processing consists of myoelectric signal segmentation, feature extraction and feature vector production [34]. In general, feature vectors are the input for pattern recognition algorithms. Feature vectors are produced by combining signal features. The combinations are empirically composed of the signal features which have demonstrated ability to provide sufficient accuracy for motion classification [34, 65, 60].

The input strongly affects the performance of pattern recognition algorithms. Technically, the accuracy of the same data set may vary depending on the use of different classifiers. It requires intensive validation to maximise the benefit of a classifier. It can be seen that improving the input for the classifiers is an alternative solution. Since data is well organised, the effort of validation for classifiers can be significantly reduced. The original data can be converted by using an individual signal feature or using a comprehensive set of signal features.

There have been a few attempts to provide criteria to maximise the benefit of signal features. Dimensionality reduction has received considerable attention since the number of recommended signal features has grown dramatically in recent decades. Dimensionality reduction can be categorised into feature selection (FS) and feature projection (FP). Implementation by feature selection is the process of preserving semantics of the MES while implementation by feature projection provides the best combination of transformed signals [37].

2.3.1 Research on feature projection and feature selection

One of the earliest studies of MES preprocessing is the research by Englehart et al. [22]. He made comparison between FP and FS, and claimed that FP using principle component analysis (PCA) outperformed FS using class separability (CS). They utilised time domain (TD) features which consisted of zero crossing, mean absolute value and trace length; and time-frequency representations (TFRs) which consisted of the short-time Fourier transform (STFT), the wavelet transform (WT), and the wavelet packet transform (WPT). When PCA and CS are applied to those

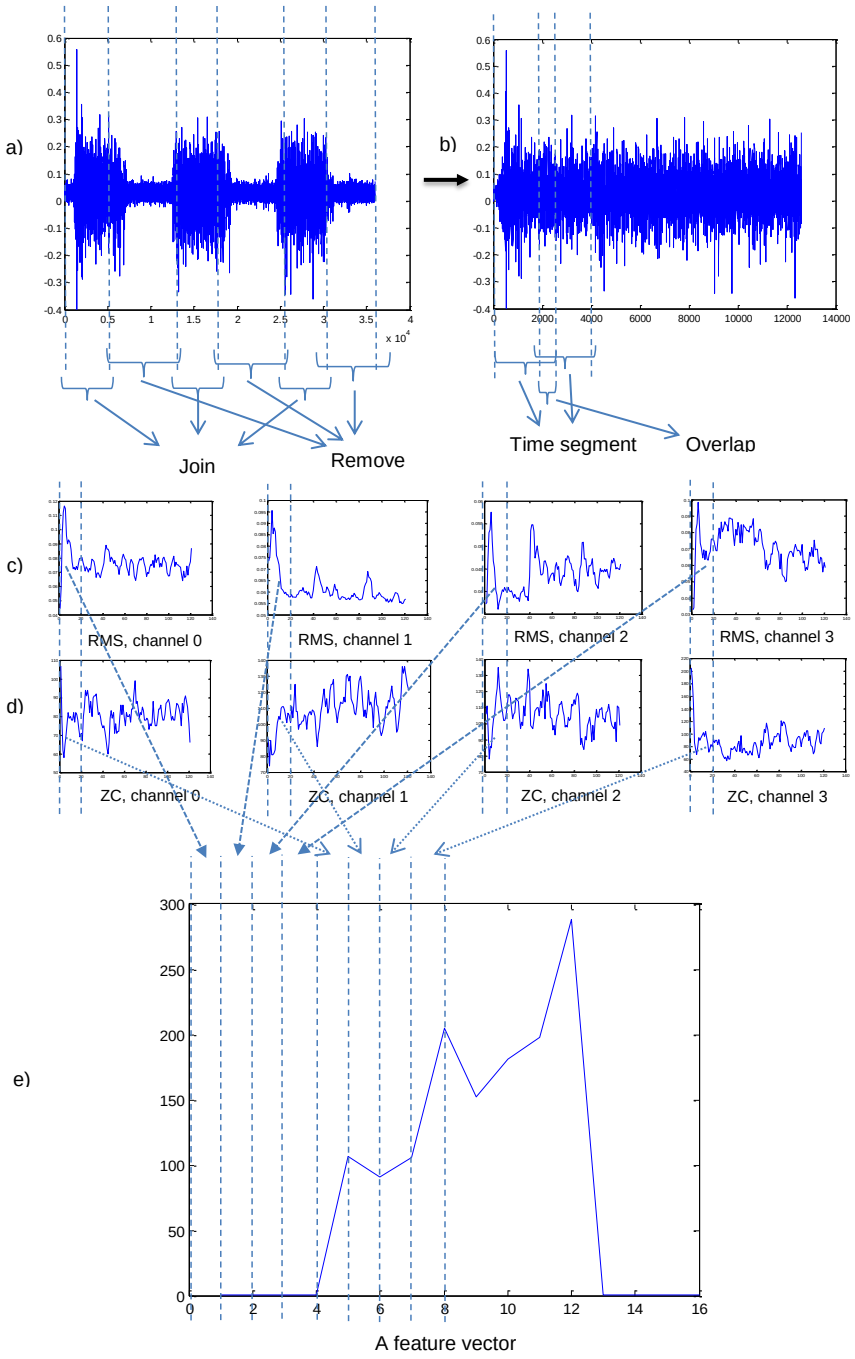
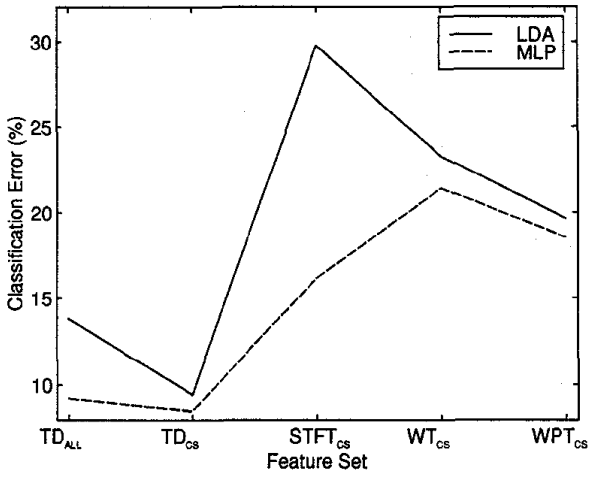
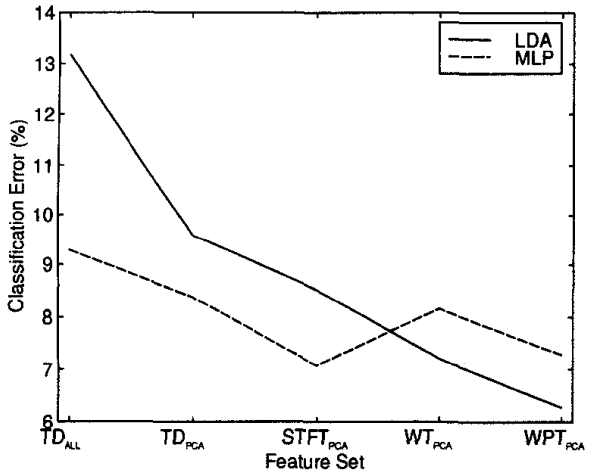


Figure 2.1: (a) The raw sEMG signal of subject 7, open hand, channel 0. (b) The pre-treated sEMG signal with 70% of contraction time. This data is segmented into 200 ms time segment (50 ms sliding time). (c) RMS signal feature computed over 121 time segments. (d) Zero crossing (ZC) signal feature computed over 121 time segments. (e) An input feature vector of 4 signal features: RMS, number of zero crossing (ZC), slope sign change (SSC) and mean of the squared absolute value (PWR) extracted from 4 channels. Each feature vector stores data points of the same time segment (200 ms) of many features in many channels.



(a) Class separability



(b) Principle component analysis

Figure 2.2: Classification error averaged across all subjects. The results show full time domain TD_{ALL}, reduced feature sets using CS (a) and using PCA (b) [22].

signal features, measurement of means of unsupervised dimensionality reduction was taken for PCA and Euclidean distance was taken for CS.

To test the accuracy of classification, they employed Linear Discriminant Analysis (LDA) and Multilayer Perceptron (MLP) methods. Fig. 2.2 shows classification errors of an entire TD feature set (TD_{ALL}), and the lower dimensional sets produced by CS and PCA. The range of classification errors of CS was from 10 to 30 and was from 6 to 14 for PCA. Overall, PCA observed the most significantly small errors. This motivated other research to focus on FP or to combine FP and FS. However, observing the errors of TD features, the errors using CS and PCA varied within a similar range, from 9 to 13 for TD_{ALL} and from 8 to 10 for TD_{CS} and TD_{PCA}. It shows that TD features did not benefit by either CS or PCA.

For the combination of FP and FS, Buchenrieder [9] suggested Guilin Hills selec-

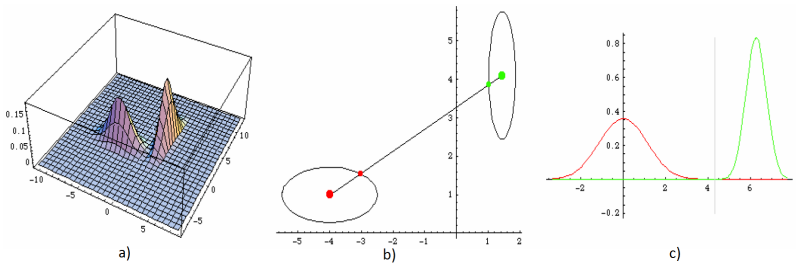


Figure 2.3: Two movement indicated by two features (a), cross sectional cut-line on the contour plot (b), baseline constructed for the combined density function in (c) [9].

RMS2	WFL1	WFL2	ZC1	ZC2	
1.3952	.258550	.0201346	.00440842	.0210817	RMS1
	.208677	.0501229	.00391864	.00391293	RMS2
		.0405795	.00352604	.02322640	WFL1
			.00335150	.00446167	WFL2
				.00453696	ZC1

Figure 2.4: Crossover distance [9].

tion method. Nine standard TD features were computed to recognise four hand movements. He proposed PCA based on FP to transform multidimensional feature space into lower dimensions, followed by FS to preserve signal features. The selection made use of the minimum crossover distance of the two features in the *two-dimensional Gaussian function*. Buchenrieder calculated the cross-sectional cut of two feature x and y

$$
\begin{aligned}
x &= d_1 \cos\varphi; \\
y &= d_2 \sin\varphi; \\
d_{cs1} &= \sqrt{\frac{d_1^2 d_2^2 \sec\varphi^2}{d_1^2 + d_2^2 \tan\varphi^2}}
\end{aligned} \tag{2.1}
$$

where d_1, d_2 were standard deviation of two features, φ was rotational angle of the hills on the contour plot. Fig. 2.3a depicts the hills, each hill represents a movement calculated by two features. When the hills was transformed to two-dimensional Gaussian function (Fig. 2.3c), a pair of features was selected corresponding to the minimum crossover distance.

Fig. 2.4 shows the crossover distance of pairwise features. In his results, the combination of ZC1-WFL2 provided the best separation to serve the proposed UniBw-Hand control system. However, the limitation of the method was that measurement could be hard when crossover became large.

2.3.2 Research on feature projection

Influenced by Englehart et al. [22], other researchers are seeking an approach using FP. A study by Chu et al. [13] combined PCA and Self-Organising Feature Maps (SOFMs). PCA was employed to simplify the structure of the fifth-order wavelet packet transform (WPT) feature and SOFMs was used to transform the reduced structure of WPT to other sub-input space with higher CS.

Chu's procedure of PCA learning can be summarised as follows. Firstly, a covariant matrix was constructed from the absolute values of WPT. Secondly, four eigenvectors corresponding to the largest values in each MES channel were selected to produce PCA projection matrix as shown in Fig. 2.5a.

After PCA, the SOFMs were applied in individual MES channels. The input of SOFMs were composed by five outputs of PCA, the output of SOFMs was a 40-by-40 two-dimensional lattice. Chu demonstrated that data processed by SOFMs improved the CS compared to processing by PCA (Fig. 2.5b).

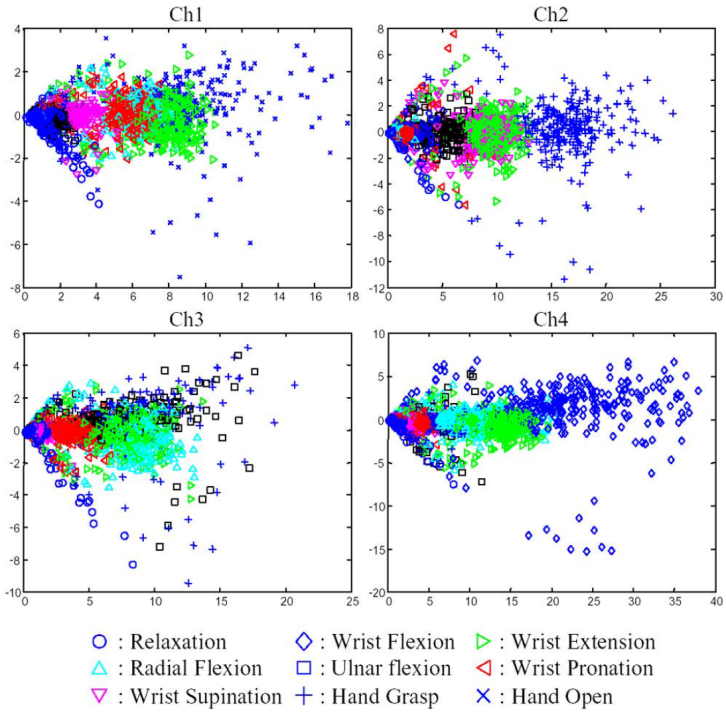
With MLP was the pattern recognition classifier, accuracy achieved by PCA + SOFMs was 97.024%, while it was 97.785% using only SOFMs and 95.759% using only PCA. The standalone SOFMs was superior to PCA + SOFMs. However, Chu argued that SOFMs required 180 ms processing time while the combination only required 5 ms. This was because SOFMs had to directly tackle raw high dimensionality feature vectors. A conclusion was that PCA + SOFMs were suitable for real-time control.

In other study, Chu et al. [14] presented Linear Discriminant Analysis (LDA) as a linear supervised projection method. WPT was used as a signal feature which is extracted to a 1024-dimensional feature. In data preprocessing, LDA reduced the linear dimensionality of WPT to eight features corresponding to eight largest eigenvalues of the covariance matrix. The projected features were applied to MLP to recognise nine hand and wrist movements. They evaluated the performance of LDA with three other methods including PCA, Non-Linear Discriminant Analysis (NLDA) and SOFM.

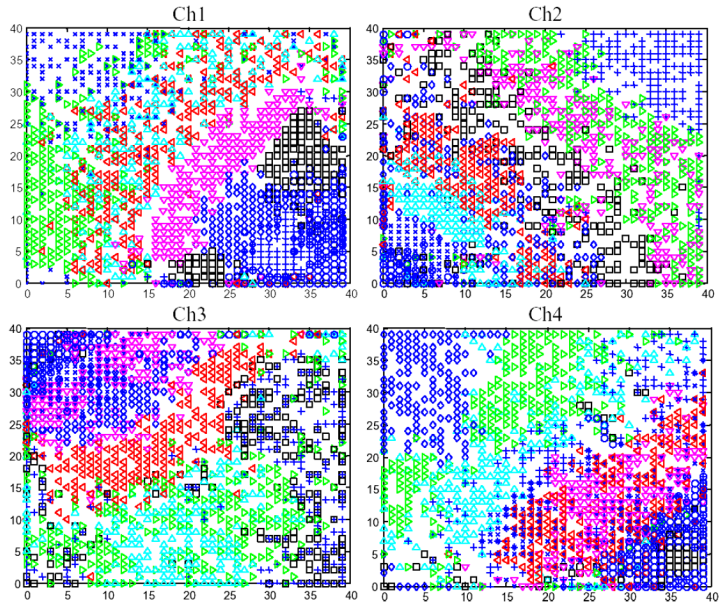
Fig. 2.6 shows reduction in features by using (a) LDA, (b) PCA, (c) NLDA, and (d) SOFM. LDA and NLDA were good differentiators of CS because they performed supervised learning. Chu et al. claimed that LDA was superior to NLDA because of the short processing time. LDA required 2 ms while NLDA needed 150 ms which exceeded the limitation time of 125 ms of the time window increment used. Comparing the accuracy, LDA achieved 97.2%; while PCA, NLDA, and SOFM achieved 94.0%, 97.3% and 95.6%, respectively. As a result, LDA demonstrated the ability to transform high dimensionality input to a sub-space of lower dimensionality.

2.3.3 Research on feature selection

There is much less research on FS compared to the research on FP. One notable study on FS was that of Khushaba et al. [41], who classified seven hand and wrist movements using six signal features. They proposed a binary Particle Swarm Optimisation (PSO) in combination with Mutual Information (MI), named as BPSOMI.



(a) PCA



(b) SOFM

Figure 2.5: The five-order WPT projected by PCA (a) and proceeded by SOFM (b) [13].

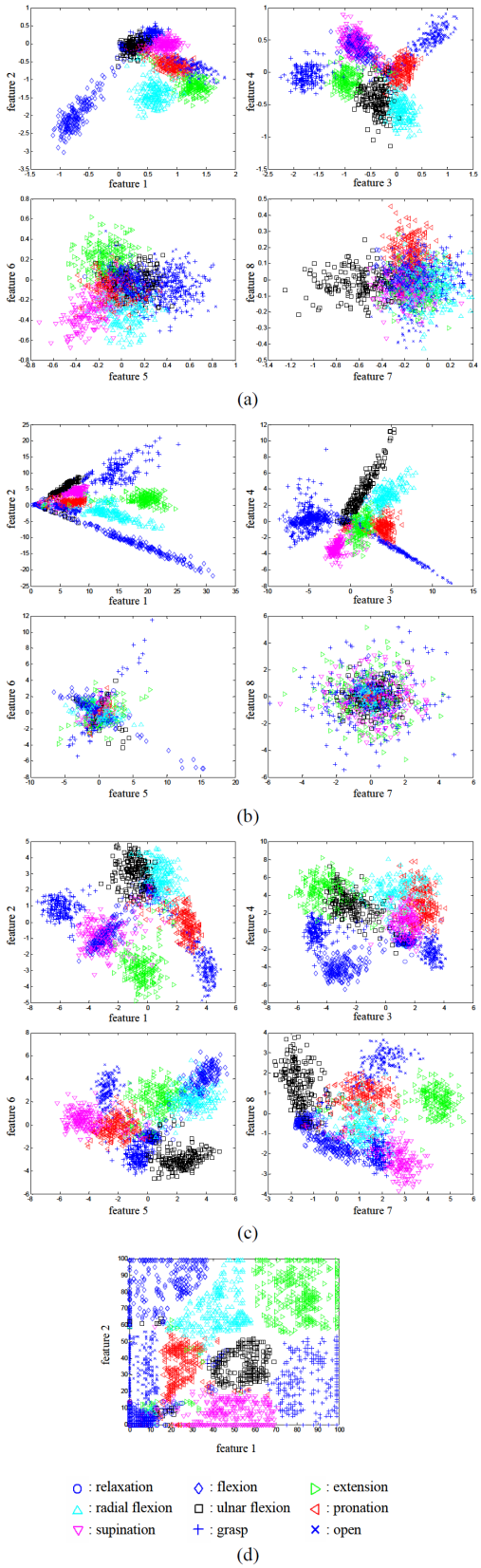


Figure 2.6: Input space of dimensionally-reduced features using (a) LDA, (b) PCA, (c) NLDA, and (d) SOFM [14].

Firstly, the binary PSO was employed to search the best sets of features and find the optimal number of features in a set. Secondly, MI was applied to measure the important of features in those sets.

The signal features used were the wavelet transform (WT) and a combined feature TDAR. The TDAR consisted of the root mean square which was a TD feature, and the autoregressive coefficient (AR). Data was treated at original state (Initial), with majority vote (MV), removal of transitional data (NT) and both majority vote and removal of transitional data (MV+NT). They compared their proposed method with the Uncorrelated Linear Discriminant Analysis (ULDA), and PCA.

Khushaba's results are in Fig. 2.7. The accuracy achieved by BPSOMI outperformed the other methods in all levels of treated data; and with both TDAR and WT. The limitation of this method is that it requires the extension of classification because signal preprocessing is attached to classification.

Most recently, Huang *et al*. [33] presented a FS method using Ant Colony Optimization (ACO). In general, ACO consists of two steps: step 1, randomly generate initial candidates by a pheromone model; step 2, adjust the pheromone values using the current candidates toward the higher solutions. The minimum redundancy maximum relevance criteria (mRMR) which is an MI approach is added between two steps to calculate the heuristic values. Huang et al. examine two types of features, the first type is a combination of 6 TD features and autoregressive, referred as TDAR; the second type is WT feature. The accuracy is calculated by MLP and implemented by majority vote. Huang et al. conducted a comparison of their FS method to the PCA which is a FP method. The accuracies achieved using ACO-mRMR on TDAR and WT are 95.45% and 96.08%, respectively; while the accuracies using PCA on TDAR and WT are 91.51% and 89.87%, respectively.

In conclusion, a majority of recent studies have focused on FP since Englehart et al. [22] concluded that FP was superior to FS. However, only a small number of TD features (three features) were involved in their study generating much lower input dimensionality. As a result, FP did not achieve high performance on TD features as opposed to the performance seen on TFD features. The classification errors of TD features also demonstrated similar ranges on both FP and FS. Due to the popularity of using TD features and FS method for pattern recognition, we will pay close attention to them. We will explore the agreement of TD features and FS method.

2.4 Summary

This chapter has reviewed the two main approaches of dimensionality reduction, namely feature projection and feature selection. The purpose is to improve the quality of the input for pattern recognition algorithms. FP has received considerable attention, while FS has received less attention in the literature. FP is a technique to provide the best combination of signal features by transforming the original vectors of signal features to a sub-space of lower dimensional vectors. FP is suitable for TSD or TFD features because they produce significantly higher dimensional input vectors. Typical TFD features are autoregressive coefficient, Short-time Fourier

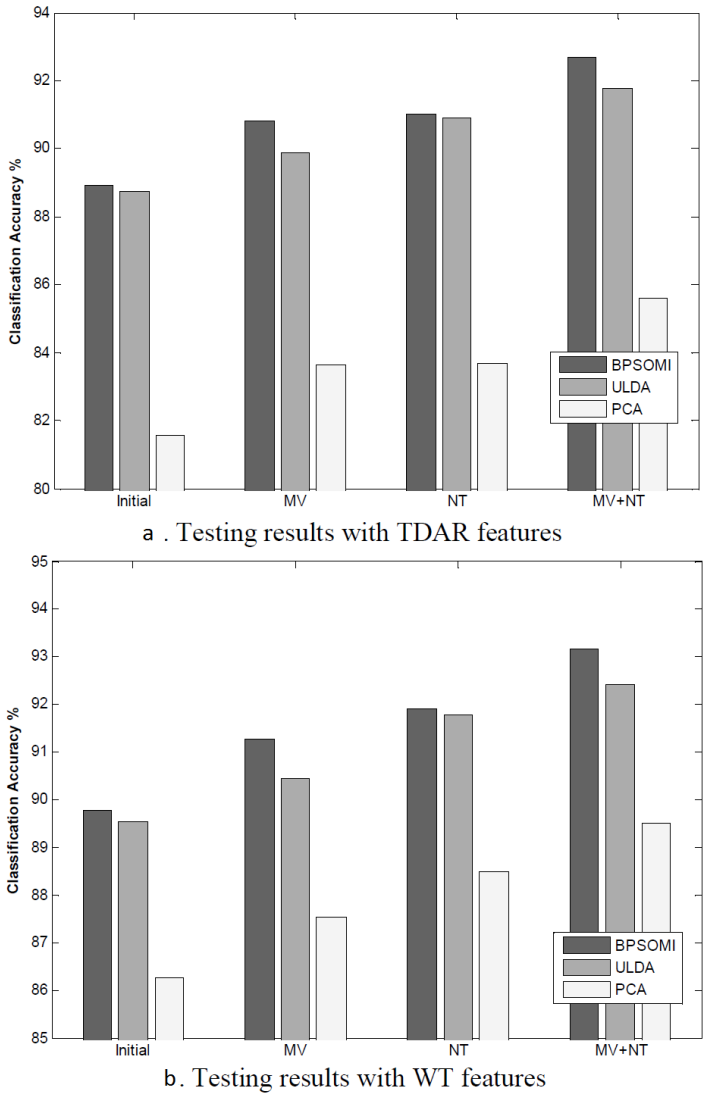


Figure 2.7: Classification accuracy achieved by using BPSOMI, ULDA and PCA across TDAR and WT feature [41].

Table 2.1: Summary of dimensionality reduction methods applied in MES features in the literature.

Author	Method	#Features	Feature type	#Movements	Movement type	Classifier	Accuracy (%)	References
Englehart *et al.*	Compare selection (CS) and projection (PCA)	6	TD, TFD	4	Hand	LDA, MLP	93.75	[22]
Buchenrieder	Combine selection (CS) and projection (PCA)	9	TD	4	Hand	MLP	-	[9]
Englehart *et al.*	Projection: PCA	1	TFD (WPT)	6	Hand	Bayesian	98	[21]
Chu *et al.*	Projection: PCA & SOFMs, LDA	1	TFD (WPT)	9	Hand and wrist	MLP	97.02, 97.20	[13, 14]
Huang *et al.*	Projection: SOFM	2	TD	8	Hand	MLP	98.75	[32]
Khushaba *et al.*	Selection: PSO	-	TD, TFD, TSD	10	Hand and wrist	MLP	97	[40]
Khushaba *et al.*	Selection: binary PSO & MI	9	TD, TFD, TSD	7	Hand and wrist	LDA	93	[41]
Huang *et al.*	Selection: ACO-mRMR	8	TD, TSD, TFD	8	Hand and wrist	MLP	96.08	[33]

transform, wavelet transform, and wavelet packet transform. To achieve a high accuracy of hand movements, the literature shows that a minimum of four MES channels are preferred for signal acquisition. A deep extraction of TFD features on each channel leads to a dramatic increase in feature vector dimension. Typical algorithms in FP are PCA, SOFM, and LDA. They convert the original feature vectors to the network's weight vectors where only significant values are selected. These values are used as the new input for pattern recognition algorithms. FS, on the other hand, is a technique of preserving semantics. It is suitable for TD features because a feature produces only one value per channel. Thus, FS preserve the most suitable signal features for the input. Since there is a limitation of available FS, for example, the remarkable methods are PSO and ACO which both use meta-heuristic search. This thesis examines other FS methods such as statistic criteria, unsupervised learning neural network and also the heuristic search. A summary of FP and FS in the literature is given in Table 2.1.

2.5 Motivation

Pattern recognition has become a potential control scheme of motorised prostheses. Several pattern recognition algorithms have been introduced and reported optimistic achievements of high accuracy. Over 95% accuracy is achieved on healthy subjects and on visual arms. Nevertheless, only 60% accuracy in real-time control reports a poor performance of these classifiers. A similar range of classification rate suggests that these algorithms have reached the optimum. This motivates us to pay attention on the input which is the process of assembling feature set. While improvement of pattern recognition algorithms requires exaggerated validation, implementation of feature set consolidates the accuracy of all algorithms without complicating them. In addition, enhancing the input sources reduces memory requirements, increases the capability to cooperate new control schemes.

TD features are used because they provide many advantages compared to other feature types. First of all, TD features are introduced for the first time in the history of contemporary prosthetic control. They show the ease of extraction by using simple mathematical functions, and can be used by many microprocessors

and microcontrollers. Second of all, TD features produce lower input dimensional vectors that do not require input reduction. Moreover, the accuracy achieved by using TD features is as high as the other types of features. The limitation of TD features is that one feature provides only one condition of the MES. Therefore, combination of multi-feature, named as a feature set, is required for multi-function control.

Selection of appropriate inputs is the first step of pattern recognition. Contemporary approach employs dimensionality reduction that can be divided into two categories namely, feature selection (FS) and feature projection (FP). Although the FS is the most appropriate method for TD features, there has not been a novel effective FS in the existing literature. We propose three approaches to generate feature sets in an attempt to achieve higher accuracy of hand and wrist movements using FS.

2.6 Research aims

The objective of this research is to find alternative sets of signal features using a minimum number of TD features to reduce the input dimensionality. Three approaches are introduced corresponding to three specific objectives:

- The first approach is a heuristic search engine which employs a genetic algorithm (GA) aiming to achieve the highest accuracy for each dataset. The GA generates several individual feature sets which are the most suitable for individual datasets.

- The second approach is application of statistical criteria which applies separately mutual information (MI), the two-sample t-test (t test) and the Mann-Whitney U test (U test). We aim to find a stable feature set suitable for all datasets. This does not attempt to achieve the highest accuracy but instead, accommodate a majority of datasets and enhance the accuracy of low quality data. This method investigates the waveforms in discrete time windows.

- The third approach also aims to find a stable feature set that can be used for several datasets. It is motivated by the unsupervised learning neural network which employs SOFM. It investigates the MES in a time series in which the SOFM searches for the similarity of signal features along the transient waveform.

Chapter 3

Genetic Algorithms for Optimising Feature Set

3.1 Introduction

In prosthetic control that uses sEMG features as inputs for pattern recognition, when too few features are extracted insufficient information is available for pattern recognition of hand and wrist movements, and when too many features are extracted, the structure of the information is divided into a larger number of cells that carry smaller data points [7]. As such, the number of inputs complicates the discussion of information in prosthetic control. Increasing size of a feature set leads to exponential growth in the dimensionality of the input space, and lower quality of training data to specify network mapping. Bellman [5] refers to this as *the curse of dimensionality*. In motorised prostheses, the dimension of the input space increases exponentially by the product of the number of signal features and the number of recording channels. Hence, using numerous input elements leads to a dramatic rise of the input space, while creating sparse data points. As a result, it does not improve the mapping representation.

In this chapter, we introduce a genetic algorithm (GA) to answer the question of how many features and which combination of features, known as feature sets, provide optimal information for pattern recognition based prosthetic control. GA is a machine learning approach that simulates natural evolution which is the process of competition, selection and reproduction. In natural evolution, a population that adapts quickly to its environment will survive and expand, while less adaptive populations will be rejected. We apply a GA as a heuristic search to randomly produce a population of feature candidates which are feature sets. The best set is reserved and reproduced, while poor sets are discarded. We measure the performance of feature sets by the accuracy hand and wrist movements. The drawback of GA is that the algorithm can be trapped in local maxima. To overcome this; we improve the fitness evaluation function, selection criteria of candidatures, mutation probability and initialisation of the feature set's population.

The initialisation of the feature set's population also provides the solution for how many features are in a set. We initialise feature sets in a variety of lengths, expecting

that the final set converges to a unique and optimal length. In addition, the best feature sets are drawn gradually through generations by replacing poor sets with better sets.

The GA is tested by Rastrigins function in a case study where an optimal solution is already known. The results show that our GA implementation is successful in finding the optimal solution of the Rastrigins function and thus can be applied to enhance the accuracy of hand and wrist movements.

This chapter explains the implementation of GA and reports our experimental results. Although the GA improves the accuracy, it involves higher computational load and increases the dimension of the input space. The results show that when accuracy reaches a certain level, increasing the length of input vectors will no longer increase the accuracy. Thus, alternative solutions are introduced to balance the cost of enhancing movement accuracy while maintaining low input dimensions.

3.2 Genetic algorithm

The conventional procedure of MES pattern recognition is illustrated in Fig. 3.1a. The input is a set of typical signal features that is selected from the literature. Since few studies focus on the performance of feature sets, challenge remains to assign the correct movement. In this study, GA was added to assemble a closed loop system where the performance of candidate sets was evaluated. GA was responsible for searching a global feature set that maximised the accuracy achieved from a dataset. After an optimal set was found, it was used in conventional pattern recognition, and GA was then removed (Fig. 3.1b).

A GA linked to two mechanisms for problem solving: encoding and evaluation. The GA encoded competitive candidates into chromosomes. Each artificial chromosome consisted of a number of genes that carried the information of which signal features were involved in a feature set, and how many of them. In other words, a chromosome represented a potential feature set. There were a number of chromosomes, referred as a population of chromosomes, providing possible solutions for improving the accuracy. In addition, crossover and mutation were the natural process to reproduce offspring chromosomes, carry part of information from previous generations and formulate new genes. We used the crossover operator and mutation operator to act like a factor indicating the ability to learn and adapt to the changes in database.

3.2.1 Encoding chromosomes

Before applying GA, chromosomes were created. A chromosome represented a feature set. In this study, each chromosome was a binary string of 21 genes encoding 21 potential signal features.

Given the 21 signal features used in a fixed order numbered from 1 to 21, a chromosome consisted of 21 genes, each gene g corresponds to signal feature f:

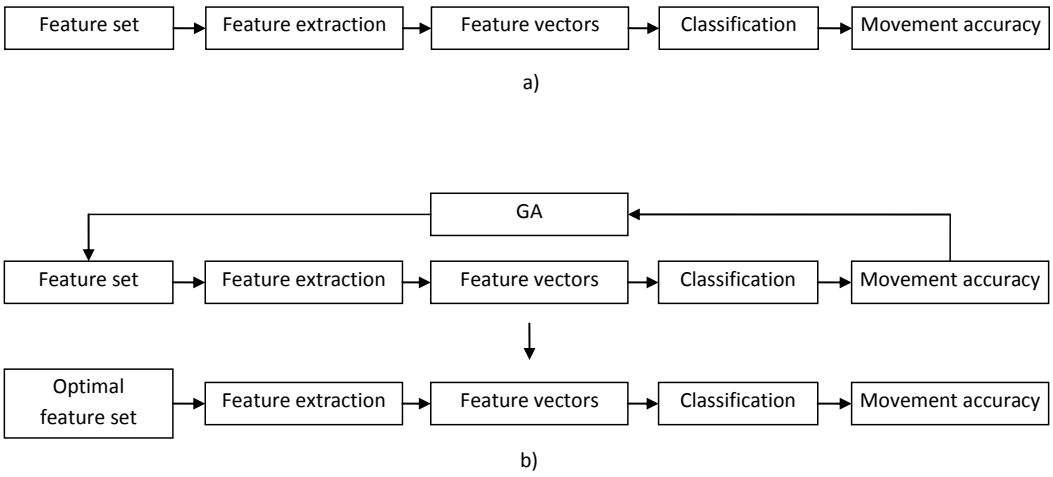


Figure 3.1: Typical procedure of prostheses classification based on patten recognition (a) and the implementation by genetic algorithm (GA) (b).

$$chrom = \{g_1, ..., g_n\} = \{f_1, ..., f_n\}$$

where n was the number of signal features. When a gene was activated, it was denoted by 1, and the corresponding signal feature engaged in a feature set. Similarly, when a gene was not activated, it was denoted by 0, and its signal feature was excluded from the feature set (Fig. 3.2).

A feature set was to generate feature vectors that were the input for a pattern recognition algorithm. In time domain, since we used four MES channels, a signal feature generated four values; they were elements of an input vector. A feature set consisted of many signal features; thus, the length of an input vector was defined by l = 4n, where n was the number of signal features in a set.

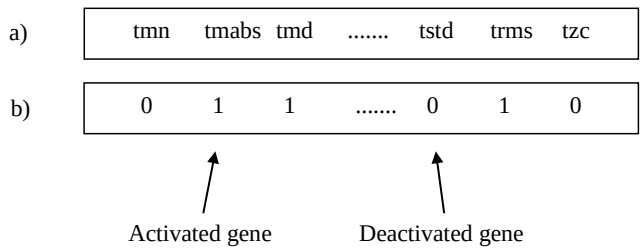


Figure 3.2: a) A set of signal features is encoded to b) an artificial chromosome. A signal feature corresponding to gene 1 comprises to the set; a signal feature corresponding to gene 0 is excluded.

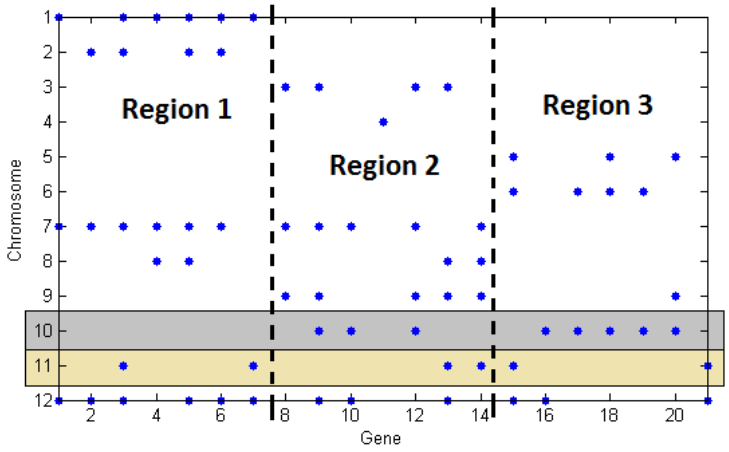


Figure 3.3: Chromosome distribution of the initial population. A chromosome is indicated by a row vector including twenty one genes. A marker indicates an activated genes, an empty space in a row indicates a deactivated gene. This figure shows chromosome 10 that has activated genes in region one 2 and 3 while chromosome 11 has activated genes in all three regions.

Four values from a signal feature were calculated in a time segment. It indicated a sample of the transient waveform. Hence, a feature vector showed an observation of the MES in a time segment.

3.2.2 Initialisation of feature set's population

To avoid the GA getting trapped in a local maximum, a high diversity of chromosomes was initialised by activating genes over the entire range of the gene's population. This allowed the GA to search in a larger region of the input space. Also, the twenty one genes considered were divided equally into three regions. The initial population of chromosomes ensured that all genes in three regions were activated. This guaranteed the contribution of signal features both in variety and quantity. Fig. 3.3 shows the initial chromosome population was equally distributed in the three regions. Chromosome 1 and 2 initialised in region 1 including genes activated in this region, chromosome 3 and 4 in region 2, chromosome 5 and 6 in region 3, chromosome 7 and 8 in region 1 and 2, chromosome 9 and 10 in region 2 and 3, chromosome 11 and 12 in all three regions. It was expected that the variety of chromosomes would converge after several generations and homogeneous chromosomes would appear at the last generation.

3.2.3 Genetic algorithm flowchart

We evaluated feature sets by measuring the fitness of individual chromosomes to select and reproduce new generations of chromosomes. Generally, the algorithm followed these steps (Fig. 3.4):

1. Presented a number of chromosomes, referred to as a population of chromosomes, to the network. The population was created randomly.
2. Defined a fitness function for classification of hand and wrist movements. The fitness function must be a pattern recognition algorithm; in this case, the artificial neural network multi-layer perceptron (MLP).
3. Run the fitness function to compute the accuracy of each individual chromosome.
4. Evaluated fitness of chromosomes. Higher accuracy represented higher fitness. Scaled the fitness by their rankings.
5. Presented the scaled fitness to the Roulette Wheel (described in section 3.2.5).
6. Selected 80% chromosomes as parent chromosomes.
7. Reproduced pairs of offspring chromosomes by using the genetic operators: crossover and mutation.
8. Preserved 20% best fit chromosomes, and replaced the rest with offspring chromosomes.
9. Placed the new population of chromosomes to the network.
10. Went to step 3, and repeated until the last generation reached.

The last generation was a terminal condition. The termination condition was commonly a defined number of iterations or generations. Because GA was a stochastic search engine, the population fitness might remain unchanged for a number of generations until a superior chromosome emerged. Therefore, it was common to stop the algorithm at a certain generation and repeated the whole process.

In the following sections, we explain the fitness evaluation, selection criteria, crossover and mutation probability.

3.2.4 Fitness evaluation

The fitness was the accuracy of hand and wrist movements, which was calculated by a fitness function. The fitness function is a MLP with resilient backpropagation. Among several training algorithms for MLP, resilient backpropagation had consistently achieved high classification rate, and had shown the ability to deal with high dimensional input vectors [39].

The MLP applied in the network consisted of three layers. The first layer was the input, with the number of input nodes matching the number of input vector's elements. The second layer was a hidden layer, with the number of hidden nodes equal to the number of input nodes. The third layer was the output layer, with the number of output nodes matching the number of movements that the MLP was required to classify. Data samples were randomly divided into training, validation and test set by the ratio of 70:15:15 for these set respectively.

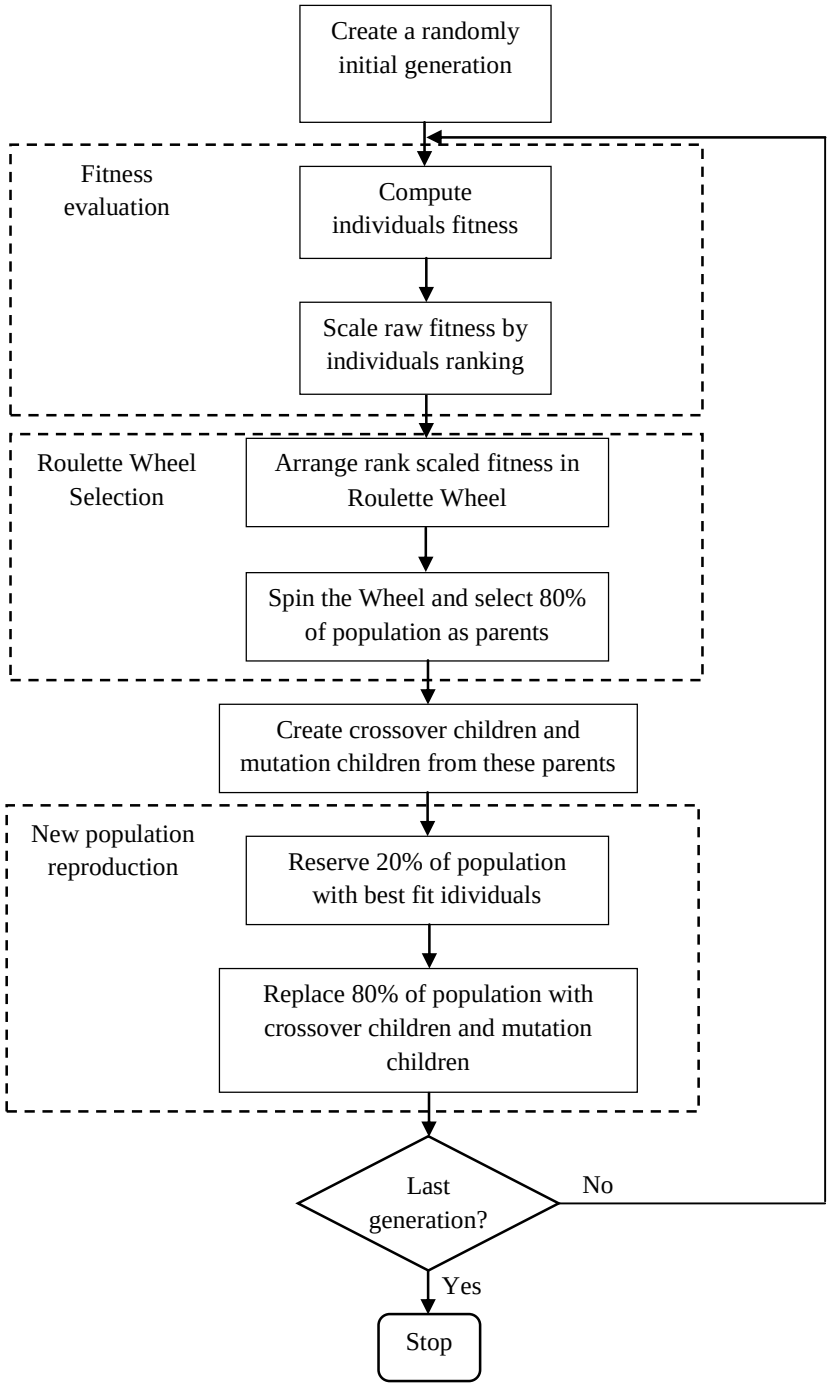


Figure 3.4: Genetic algorithms flowchart for finding optimal feature sets.

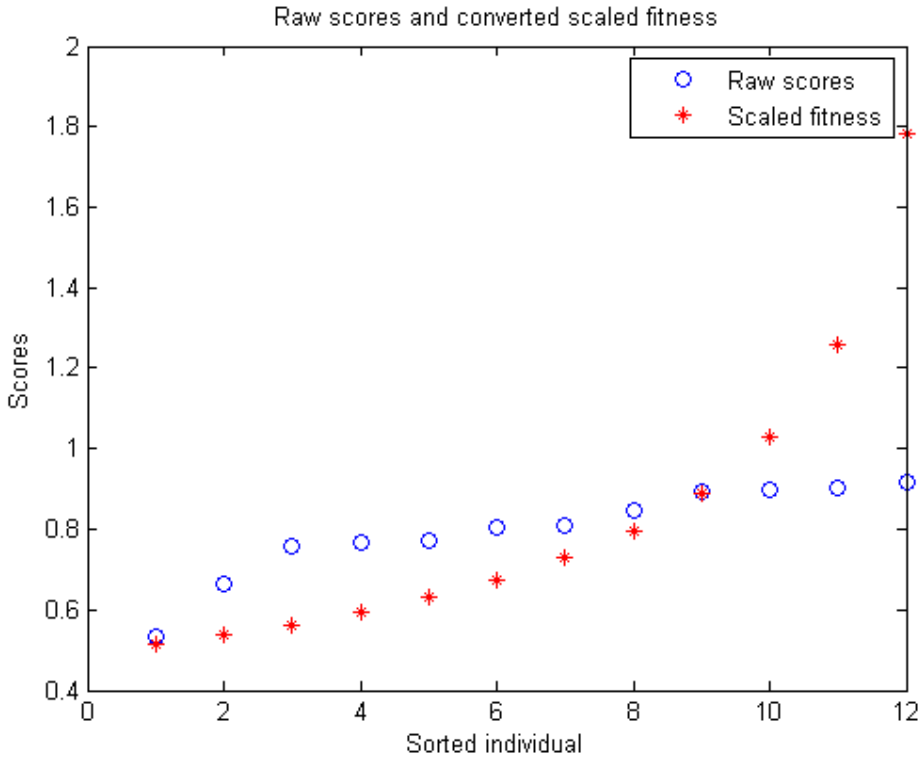


Figure 3.5: Raw scores and converted scaled fitness.

Each chromosome in a population was one-by-one fed to the MLP to calculate its fitness, which was the average accuracy of all accuracies of movements. The fitness was measured by a Fitness Rank Scaling (FRS) function. The raw fitness scores are converted into scaled fitness by the FRS. Given chromosome i, its fitness is

$$fitness_i = \frac{k}{\sqrt{rank_i}} \tag{3.1}$$

where k was the scaling factor proportional to the number of selected parent chromosomes,

$$k = \frac{\#parents}{\sum_{i=1}^{n} \sqrt{i}}$$

where n was the size of the population and $rank_i$ was the ranking of chromosome i in descending order.

The scaled fitness was helpful in removing the effect of dissemination of the raw scores, generated by the fitness function. Fig. 3.5 shows the raw scores converted to the scaled fitness of a typical population of size 12 sorted in ascending order. Because the rank scaled fitness depended only on the ranking of individuals, it would be similar for any population of 12 individuals. This guaranteed a steady range of scaled values, preventing the GA from searching in other input spaces and converging quickly within a population gene pool.

3.2.5 Roulette wheel selection

The roulette wheel was used as a selection criterion which randomly selected chromosomes for reproduction. It relied on a fixed number of parent chromosomes, size of the population and chromosome fitness. It worked on the concept that higher fitness chromosome would be more likely to be selected. Given chromosome i corresponding to fitness f_i and a population of size n, its probability to be selected was

$$p_i = \frac{f_i}{\sum_{j=1}^{n} f_j}. \tag{3.2}$$

Rotated the wheel by *#parents* times corresponding to the number of intended parents, the implementation of the selection technique was executed by a cumulative distribution function (CDF) defined by

$$F_{chrom}(x) = p(chrom \leq x). \tag{3.3}$$

A chromosome was chosen if its probability was less than or equal to x, the probability that chrom lied in the interval $(a, b]$ where a and b were the cumulative sum of previous *chrom* fitness and this *chrom* fitness, therefore

$$p(a < chrom \leq b) = F_{chrom}(b) - F_{chrom}(a). \tag{3.4}$$

A uniform random number from the range of $[0, 1)$ was chosen and multiplied by the *chrom* fitness to generate x. This number representd a ball in the roulette wheel, if the ball (the random uniform number) landed between a and b or equal to b, the *chrom* was selected.

3.2.6 Crossover operator

In nature, crossover is the exchange of genes between a pair of chromosomes that results in genetic recombination. The offspring chromosomes inherit the genetic material of the parent chromosomes, that results a new arrangement in the form of genes known as alleles which allows more options for natural selection. This leads to the complement of the previous generation. In genetic algorithms, the crossover operator acts in the same way as the chromosomal crossover. It varies the genetic material of the current generation and passes it to the next generation.

There are several crossover techniques; we used a fundamental technique which was one-point crossover. It happened on at a single point on both parent chromosomes. Firstly, the crossover operator randomly selected a point at a position of the two parent chromosomes selected by the roulette wheel. The operator would break the chromosome at this position, known as a crossover point. A part of chromosomes was exchanged from this point to the end of chromosomes resulting two new offsprings (Fig. 3.6). In the case of no crossover, the chromosome cloned itself, then the offsprings were copied exactly the same as its parent. The probability of

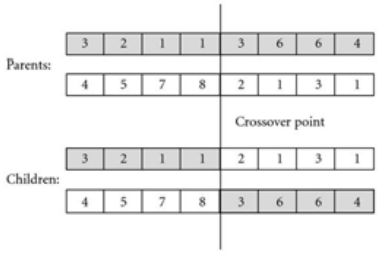


Figure 3.6: One-point crossover.

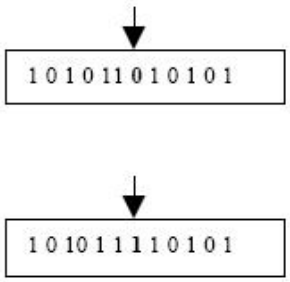


Figure 3.7: A bit string mutation.

crossover was vital to generate a good result. In our experiments, it was shown that 0.9 was practical for exceptional accuracy. It substantiates the recommendation of probability between 0.7 and 0.9 in the literature [69].

3.2.7 Mutation operator

Mutation causes a permanent change of the DNA sequence that make up a gene. This occurs due to chemical mutagens or radiation; thus, the probability of mutation is usually very small but can result in a significant improvement in the performance. The role of mutation aims to escape the trap of local optimum through the search process due to crossover. The sequence of this condition leads to a homogeneous chromosome generation, thus stagnate the fitness. The solution or search algorithm, hence, stops at a local optimum which processes no further improvement. Mutation provides a random search which is helpful to avoid this consequence.

A mutation can occur in several types. We used the most common type which was transition or bit string mutation that affects on the structure of a chromosome. It happened at some points in a chromosome that changed from a single bit (0 or 1) to another bit (1 or 0). The mutation operator randomly chose a gene in a chromosome and flipped it over (Fig. 3.7). Any gene in a chromosome could be mutated with a probability. The probability was held very small in the range (0.001, 0.01) [69], in this case $p_m = 0.005$.

3.2.8 New population reproduction

After experiencing the above process, new chromosomes were reproduced and utilised for decoding feature sets. The fitness function evaluated these new sets, updated the results and created new population of chromosomes. The new population was created by reserving 20% the best fit chromosomes from the old population and replacing the others by the new chromosomes. In the next generation, the procedure was replicated, starting from evaluating the fitness of the entire population of chromosomes. The GA was terminated when the last generation reached.

3.2.9 Case study

This section shows an example of finding the global minimum of the Rastrigin function, to test the performance of the GA. The Rastrigin function was defined by

$$f(x) = An + \sum_{i=1}^{n} \left[x_i^2 - Acos(2\pi x_i) \right] \tag{3.5}$$

where $A = 10$, n was the number of variables and $x_i \in [-5.12, 5.12]$. The function had a global minimum at $x = 0$ where $f(x) = 0$. The function was used as a fitness function of the GA under with two independent variables format

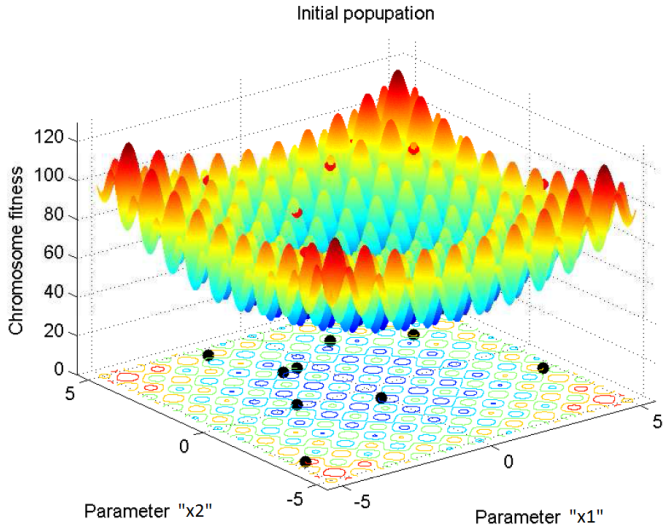
$$f(x) = 20 + x_1^2 + x_2^2 - 10(cos2\pi x_1 + cos2\pi x_2). \tag{3.6}$$

Twenty one signal features were converted to two real numbers in the interval of-5.12 to 5.12, used as the input variables. A chromosome with a length of 21 genes corresponding for 21 signals features were converted to a binary string with a length of 8, and converted from binary to real number. The fitness ranking scale converter, Roulette Wheel Selection, Crossover and Mutation operator were reserved. The performance of GA was indicated in Fig.3.8. The black dots on the contour plots in Fig.3.8 represent the chromosomes or the individuals of a population of size 10 indicated for 10 sets of the input variables of the Rastrigin function. At the initial population, the individuals were distributed widely over the whole range of the input space (from -5.12 to 5.12). After 3000 generations, ten individuals became homogeneous and converted approximately to point (0,0) which was the global minimum of the Rastrigin function.

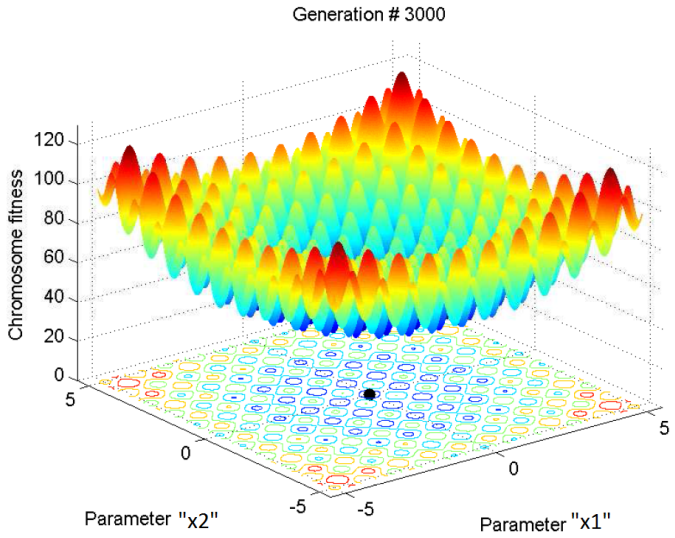
The success of the GA on the Rastrigin function demonstrated that the GA was able to escape the trap of a local optimum. In section 3.3, we present the experimental results of using the GA to find an optimal feature set for individual dataset.

3.2.10 MES data

We used pre-recorded data as described by Ortiz-Catalan et al. [59]. MES data source was collected from 20 healthy subjects of both genders, non-amputee. A total MES dataset of 10 hands and wrist movements was recorded on each subject,



(a) Initial population



(b) Final population

Figure 3.8: Testing GAs by minimising Rastrigin's function. The randomly initial (a) and final population (b) of chromosomes after 3000 generations.

Table 3.1: Signal features were encoded to genes of an artificial chromosome. A gene is represented by g_x, it can receive a value either 1 or 0.

Signal feature	Gene number	Chromosome
tmn	1	g_1
tmabs	2	g_2
tmd	3	...
tstd	4	...
tvar	5	g_x
twl	6	...
trms	7	...
tzc	8	...
tpks	9	...
tmpks	10	...
tmvel	11	...
tslpch1	12	...
tslpch2	13	...
tpwr	14	...
tcr	15	...
tcv	16	...
tdam	17	...
tfd	18	...
tmfl	19	...
tfdh	20	...
tren	21	g_{21}

including hand open (OH), hand close (CH), hand pronation (PN), hand supination (SN), wrist flexion (FH), wrist extension (EH), side grip (SG), fine grip (FG), thumb up (AG), and pointing (PT). Additionally, rest movement was considered by accumulating rest MES of other 10 recorded movements. We analysed and classified 11 movements (10 movements and rest). The data acquisition system was described by Ortiz-Catalan et al. [59]. In summary, it consisted of 4 pairs of bipolar electrodes, referred as 4 MES channels, placed around the forearm of the subject. The first pair was attached to the extensor carpi ulnaris, and the other pairs were placed equally spaced around the forearm. The twenty one signal features were encoded to genes of a chromosome as shown in table 3.1.

3.2.11 Signal feature computation

The signal features used in this study could be found in the literature. A feature vector was produced by computing signal features within a time segment. There were 122 feature vectors corresponding to 122 time segments in which 200 ms each segment and 50 ms time increment. Given $x_i(k)$ was the value of data at time k in segment i, with n data points, the signal feature in a time segment was computed as follow:

1. Mean value (tmn)
   The average value of data in time segment i

$$tmn_i = \frac{1}{n}\sum_{k=1}^{n} x_i(k).$$

2. Mean absolute value (tmabs)
   The average value of absolute data in time segment i

$$tmabs_i = \frac{1}{n}\sum_{k=1}^{n} |x_i(k)|.$$

3. Median (tmd)
   The median value of data in time segment i.

4. Standard deviation (tstd)
   The standard deviation of time segment i

$$std_i = \left(\frac{1}{n-1}\sum_{k=1}^{n}(x_i(k) - \overline{x_i})^2\right)^{\frac{1}{2}}$$

   where

$$\overline{x_i} = \frac{1}{n}\sum_{k=1}^{n} x_i(k).$$

5. Variance (tvar)
   The variance of data in time segment i

$$tvar_i = std_i^2.$$

6. Waveform length (twl) [34]
   The amplitude difference of time segment i between two consecutive data points k and $k+1$

$$twl_i = \sum_{k=1}^{n-1} \left(x_i(k) - x_i(k+1)\right).$$

7. Root mean square (trms)
   The RMS value of data in time segment i

$$trms_i = \left(\frac{1}{n} \sum_{k=1}^{n} |x(k)|^2 \right)^{\frac{1}{2}}.$$

8. Zero crossing (tzc)
   The number of times signal crosses zero amplitude. It represented a signal frequency perspective in time domain (TD)[34]

$$tzc_i = \sum_{k=1}^{n-1} f(k)$$

   with $f(k) = 1$ if $x_i(k)x_i(k+1) < 0$ and $|x_i(k) - x_i(k+1)| > x_{th}$, x_{th} was a threshold value, $x_{th} = tmabs$.

9. The number of peak values that over RMS (tpks)

$$tpks_i = \sum_{k=2}^{n-1} f(k)$$

   with $f(x) = 1$ if $(x_i(k) - x_i(k-1))(x_i(k) - x_i(k+1)) > 0$ and $|x_i(k)| > RMS$.

10. Mean of peaks over RMS (tmpks)
    Similar to *tpks*, returned the mean value of *tpks*.

11. Mean of the difference of peaks over RMS or the velocity of peaks (tmvel)

$$tmvel_i = \frac{1}{n} \sum_{k=1}^{n} \left| pks\left(x_i(k)\right) - pks\left(x_i(k-1)\right) \right|.$$

12. Slope sign change using tmvel value as a threshold (tslpch1)
    The number of times the slope of waveform changed sign. It represented a signal frequency measurement in TD.

$$ssc_i = \sum_{k=2}^{n-1} f(k)$$

    with $f(k) = 1$ if $(x_i(k) - x_i(k-1))(x_i(k) - x_i(k+1)) > x_{th}$ and $f(k) = 0$ otherwise where $x_{th} = tmvel$.

13. Slope sign change (tslpch2)
    The number of times the slope of waveform changed sign. It represented a signal frequency measurement in TD [34].

$$ssc_i = \sum_{k=2}^{n-1} f(k)$$

    with $f(k) = 1$ if $(x_i(k) - x_i(k-1))(x_i(k) - x_i(k+1)) > x_{th}$ and $f(k) = 0$ otherwise where $x_{th} = tmabs$.

14. Power of waveform (tpwr)

$$tpwr_i = \frac{1}{n}\sum_{k=1}^{n}|x(k)|^2 = rms_i^2.$$

15. Correlation of data between channels (tcr)

$$tcr_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

with $cov(x,y)$ was the covariance of data in two channels x and y, and σ_x, σ_y were the standard deviation where

$$cov(x,y) = \frac{1}{n-1}\sum_{k=1}^{n}\left(x(k)-\overline{x}\right)\left(y(k)-\overline{y}\right)$$

and

$$\sigma_x = \left(\frac{1}{n-1}\sum_{k=1}^{n}(x(k)-\overline{x})^2\right)^{\frac{1}{1}}$$

and

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n}x(k)$$

The correlation was written

$$tcr(x,y)_i = \frac{\sum_{k=1}^{n}(x_i(k)-\overline{x_i})(y_i(k)-\overline{y_i})}{\sqrt{\sum_{k=1}^{n}(x_i(k)-\overline{x_i})^2(y_i(k)-\overline{y_i})^2}}.$$

16. Covariance of data between channels (tcv)

$$tcv(x,y)_i = cov(x,y) = \frac{1}{n-1}\sum_{k=1}^{n}(x_i(k)-\overline{x_i})(y_i(k)-\overline{y_i}).$$

17. Mean of the absolute difference (tdam)

$$tdam_i = \frac{1}{n-1}\sum_{k=2}^{n-1}\left|x_i(k)-x_i(k-1)\right|.$$

18. Fractal dimension using Katz's algorithm (tfd)

$$tfd_i = \frac{log_{10}(n)}{log_{10}\left(\frac{d}{L}\right)+log_{10}(n)}$$

where L was the sum of Euclidean distances between consecutive points, d was the distance between the first point and the point that provided a farthest distances.

19. Maximum fractal length (tmfl) [3]

$$L_m(k) = \frac{\left(\sum_{i=1}^{\frac{n-m}{k}} |X(m+ik) - X(m+(i-1)k)|\right)\frac{n-1}{\lfloor\frac{n-1}{k}\rfloor k}}{k}.$$

20. Fractal dimension according to Higuchi (tfdh)
Fractal dimension according to Higuchi computed by drawing a straight line and taking its slope [29].

$$tfdh_i = \frac{\Delta L_m}{|\Delta k|}.$$

21. Rough entropy per channel (tren) [79, 47, 62]

$$tren_i = -\sum_{j=1}^{m} \frac{|R_j|}{|U|} log_2 \frac{1}{|R_j|}$$

where $|U|$ was the universe size, or in this application, the sample size ($|U| = n$); $|R_j|/|U|$ was the probability of unique element R_j; m was the number of unique elements in the universe U; $1/|R_j|$ was the probability of one of the value in set R_j.

3.3 Results

The GA searched for an optimal feature set on an individual subject. The average accuracy of 11 movements was reported. There was a competition within the 12 feature set in a generation. It was found that the accuracy did not rise after 200 generations. We also observed the variation of feature sets, referred to as chromosomes, from the initial population to the final population. As expected, the initial distribution of chromosomes varied in 3 regions, accommodating the variation of length of feature set. There were 21 features in our experiment potentially initialising a feature set of minimum 1 feature and maximum 21 features. We observed in the experiment that GA randomly initialised smallest feature sets of 1 feature and largest feature sets of 10 features. Ultimately, all feature sets became homogeneous in the final population.

Following is an example of applying the GA on subject 17. Fig. 3.9 shows the accuracy of the initial generation of chromosomes and the chance in the last generation (Fig. 3.9a and 3.9b). While there was only 1 chromosome generating the highest accuracy at first, indicated by the green column, there were several chromosomes achieving maximum accuracy at the end. The variation of accuracy can be explained by the inconsistency of chromosomes in the initial population (Fig. 3.9c). Some chromosomes included a few activated genes, indicated by blue dots, providing insufficient information for classification. It is highlighted that the first region of genes which consisted of amplitude signal features was inferior to the last region of genes which were fractal length signal features. The first region accuracy was above 85% while the last region achieved approximately 95% accuracy. The accuracy attained was closer to 100% when homogeneous chromosomes transpired at the final generation. It can be seen that in order to achieve the highest accuracy,

Table 3.2: Results of optimal sets acquired by GA, set 3 and a literature set on average accuracy of 11 movements on 20 subjects, Wilcoxon p-value indicates relationships between accuracies of each of new sets with the literature set.

	Computational time (s)	acc (%) ±SD	Wilcoxon p-value
Optimal sets [a]	69.070 ±32.193	92.63 ±0.037	0.0057
Set 3 [b]	1.005 ±0.209	88.81 ±0.062	0.0228
Literature set [c]	0.332 ±0.097	90.06 ±0.060	-

[a] acquired by GA for individuals
[b] *trms, tpks, tpwr, tcv*
[c] *tmabs, twl, tzc, tslpch2*

many chromosomes in three regions were required (Fig. 3.9d). The best chromosome consisted of 11 genes. In other words, the optimal feature set committed 11 TD features to classification.

Fig. 3.10 summaries the performance graph of accuracy over generations on subject 17. The best accuracy of 100% was possible to achieve but the average accuracy of all chromosomes remained at 99.5% after generation 20.

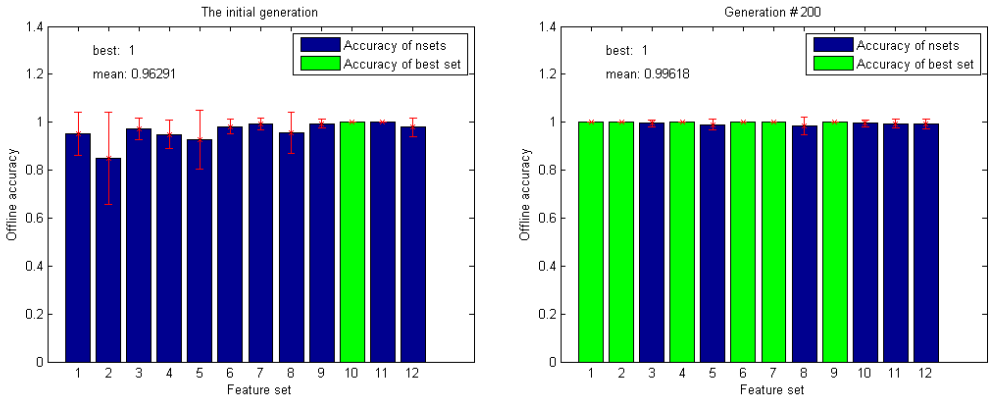
In general, we found 20 optimal feature sets for 20 subjects, more than 10 TD features in a feature set, indicated by a column in Fig. 3.11, was a demand for the highest accuracy. Similar to an example on subject 17, three regions of TD features guaranteed the best performance of classification.

Since the number of features expanded dramatically (from 4 in the literature to more than 10 found by GA), the input dimensionality increased 4 times this. Our purpose was achieving a concise feature set for each subject while retaining high accuracy. It was essential to reduce the size of the above feature sets because many researchers have shown that increase of accuracy does not accompany extension of the feature set [40, 33].

Subsequently, we searched for a common feature set for all subjects. We presented a histogram of signal features that occurred in all feature sets across subjects, as shown in Fig. 3.12. There were 4 out of 21 features frequently presented, including feature number 7, 9, 14 and 16 which were the RMS values (trms), number of peaks over RMS value (tpks), power (tpwr), and covariance of data between channels (tcv). They were named as set 3.

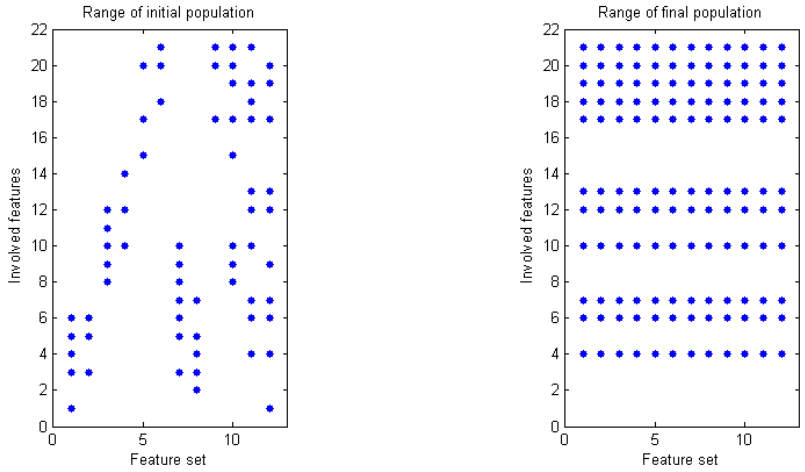
We measured the accuracy of set 3 and compared with the accuracy of optimal sets previously acquired by the GA. A literature set found in several existing studies [34, 59, 53, 60], that recommended four TD features: tmabs, twl, tzc, tslpch2; was also evaluated.

Table 3.2 compares the average accuracies of optimal sets acquired by GA, set 3, and a literature set; pattern recognition algorithm was the MLP. The accuracy was calculated over 20 subjects to classify 11 hand and wrist movements. Optimal sets outperformed the literature set, demonstrated by $p < 0.05$. However, time taken to achieve higher accuracy was significantly longer than that of the literature set,



(a) Initial accuracy

(b) Final accuracy



(c) Initial popolulation

(d) Final popolulation

Figure 3.9: Accuracy of twelve feature sets searched by GA on subject 17. The initial random sets (c) provide a variety of accuracy (a), homogeneous final sets (d) provide equivalent accuracy (b).

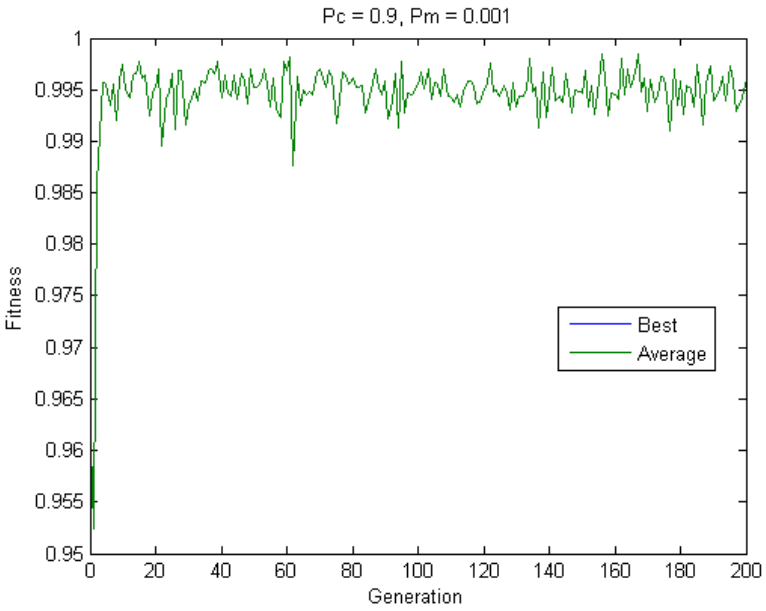


Figure 3.10: The performance graph of accuracy through generations on subject 17.

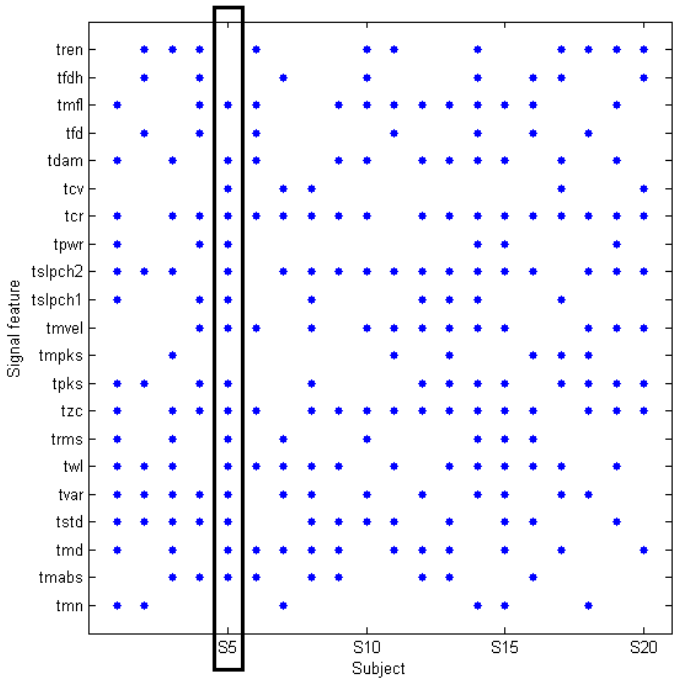


Figure 3.11: Feature sets acquired by GA on each subject. A marker in a column (e.g column in S5) indicates the appearance of a signal feature needed in a subject.

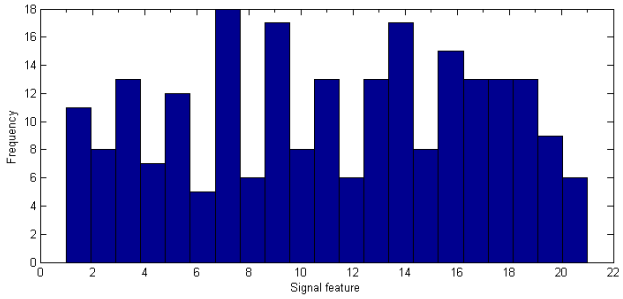


Figure 3.12: Histogram of signal features acquired by GA over 20 subjects.

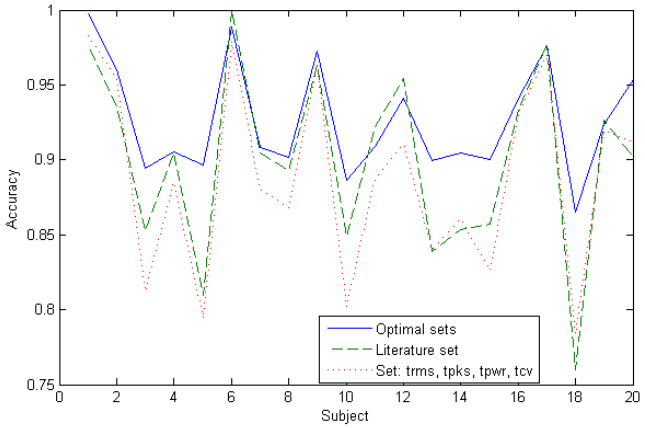


Figure 3.13: Average accuracy of the optimal feature sets for an individual subject acquired by GA, by the literature set, and by a common feature set gathered on all subjects.

over 69 seconds for the optimal sets compared to 0.332 second for the literature set. This was because there were more than 10 signal features in each set of optimal sets, compared to 4 signal features in the literature set. This was a computational burden on the MLP. Replacement set 3 significantly reduced computational time from 69 seconds to 1 second; it, however, did not achieve accuracy as high as the literature set.

In details, fig. 3.13 shows the average accuracies of these sets over all subjects. It is noticed that set 3 appeared at the bottom. Consistently, the optimal sets outperformed the other sets, and particularly improved the accuracy of subjects providing low quality data such as subject 3, 5, 10. The optimal sets achieved at least 90% accuracy on these subjects while the other sets achieved around 80% to 85% accuracy. The second best feature set was the literature set which achieved higher accuracy than set 3 in most of subjects. It is noticed that the accuracy reached the same peak for all sets at some subjects such as subject 6, 9, 17. It is seen that a small set of 4 signal features was as good as a larger set of more than 10 features.

In brief, set 3 was unsuccessful in an attempt to replace optimal sets assembled by the GA. The optimal sets for individual subjects accomplished higher accuracy, however they extended the input dimensionality. In the next chapters, we introduce alternative methods for feature selection.

3.4 Conclusion

This chapter presents a genetic algorithm to explore the individually optimal set of signal features for a single subject. The GA is added to a normal pattern recognition procedure to construct a close loop system that evaluates the accuracy produced by the feature set and, thus, makes adjustments. The GA is removed when an optimal feature set for a subject is found. Several optimal feature sets were found corresponding to several subjects. For classification, MLP-resilient backpropagation is the fitness function as well as the pattern recognition classifier. Experimental results show that the GA successfully found optimal feature sets achieving higher accuracy.

Although our proposed method improves the accuracy, it leads to an extension of the input dimensionality. Fewer classifiers are able to adapt to a set of more than ten features, particularly with classifiers that require complicated computation. To accommodate the GA, we have developed a further step to reduce the size of optimal feature sets.

Common signal features assemble a global set, named as set 3, which consists of four signal features. Unfortunately, set 3 shows poor performance, as opposed to the optimal sets and a literature set. In the next chapters, we introduce more powerful methods in an attempt to provide alternative search engines for the solution of feature selection.

The major contribution of this chapter is an investigation of GA as an alternative method in the field of heuristic search. It is a conclusion that GA can be a solution for FS if the size of a feature set retains sufficiently small.

Chapter 4

Class relevant coefficient criteria

4.1 Introduction

This chapter describes three types of search engines for feature selection (FS) to find an optimal feature set. In contract to the GA, this method searches only one feature set capable of several subjects. It is motivated by a relationship of the movement classes and data distribution. The first engine is *mutual information* (MI), aiming to measure the probability of join distribution $p(x, y)$ to the products of their individual probability $p(x)p(y)$. The second and the third search engine employ statistic tests. The second search engine applies the *two-sample t-test* (t test) in the purpose of measuring the variances and means of two populations of data extracted from a signal feature that separates a class of movement from other classes. The third search engine implements the *Mann-Whitney U test* (U test) in exchange for t test by an assumption of non-normally distributed data. This chapter also presents three pattern recognition algorithms used for evaluation of the feature sets including Multilayer Perceptron (MLP), Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM).

4.2 Mutual Information with Max-Relevance and Min-Redundancy

The Mutual Information (MI) used Max-Relevance and Min-Redundancy to calculate an index of a signal feature with a movement class. Signal features with the highest indexes were selected. The calculation was presented as follows.

Given two random signal features x_i and x_j as the inputs, their mutual information was defined in terms of their probabilistic density functions

$$I(x_i; x_j) = \iint p(x_i, x_j) log \frac{p(x_i, x_j)}{p(x_i)p(x_j)} dx_i dx_j. \tag{4.1}$$

Mutual information represented the sharing of information between variables x_i

and x_j. If these variables were independent, $I(x_i; x_j) = 0$. According to Peng [64], Max- Relevance was applied to a signal feature x and class c to select feature x_i which acquired the largest mutual information $I(x_i, c)$ to the target class c, representing the largest dependency on the target class. The Max-Relevance criterion provided good signal features with the most information of the target movement c. Given a feature set $M = \{x_i : i = 1, ..., m\}$ with membership feature x_i and number of features m in the set, Max-Relevance searched set M^* to satisfied the condition

$$D(M^*, c) = \arg\max_j D(M_j, c),$$
$$D = \frac{1}{|M|} \sum_{x_i \in M} I(x_i; c). \tag{4.2}$$

Selection according to Max-Relevance could result in high redundancy because the dependency among these features and the target class could be large. Therefore, the criterion of Min-Redundancy was added [64]

$$R(M^*) = \arg\min_j R(M_j),$$
$$R = \frac{1}{|M|^2} \sum_{x_i, x_j \in M} I(x_i; x_j). \tag{4.3}$$

The definition of *minimal-redundancy-maximal-relevance* (MRMR) [64] was to combine them following the simple optimization

$$\Phi(D, R) = \arg\max (D - R). \tag{4.4}$$

The MRMR condition measured signal features one-by-one. The first-selected signal feature corresponded to the largest $\Phi(D, R)$ within the entire set M, the second-selected signal feature was the largest $\Phi(D, R)$ within set $M - 1$, and the procedure was replicated until the required number of features exceeded.

4.3 Two-sample t-test

The two-sample t-test (t test) and Mann-Whiteney U test (U test) were statistical tests applied in this area based on the idea of evaluating the responses of individual features on movement classes which separated two label classes one at a time. The t test was a normal distribution test, also known as Welch's t-test working under the following conditions:

4.3.1 Assumption

1. The response of MES data in a class was independent of those in the other classes.

2. Observations were normally distributed.

Individually, one class of movement was assigned against the others, the assigned class was denoted by 1 and the other denoted by 0. Given a signal feature with 2 groups of samples corresponding to 2 classes, sample 1 for class 1 was, $x_1 = \{x_{11}, x_{12}, .., x_{1n1}\}$ and sample 2 for class 0 was, $x_0 = \{x_{01}, x_{02}, .., x_{0n0}\}$. We computed the test statistic t-value.

4.3.2 Computation

1. Computed the test statistic t-value:

$$t = \frac{\overline{x_1} - \overline{x_0}}{\sqrt{\frac{s_1^2}{n1} + \frac{s_0^2}{n0}}} \tag{4.5}$$

   where $\overline{x_1}$ and $\overline{x_0}$ were the mean values of x_1 and x_0, s_1 and s_0 were the standard deviations of two samples.

2. Compared the calculated t-value to the critical t-value from the table of t distribution under the null hypothesis. Reject if

$$t > t_{critical}.$$

4.4 MannWhitney U test

The Mann-Whiteney U test (U test) also known as Wilcoxon rank-sum test was more efficient than the t-test on non-normal distributions. To apply this test, we made the following assumptions about the MES data:

4.4.1 Assumption

1. All samples from any two classes were independent.
2. Data were measured at a continuous level. i.e. differentiated data by MES voltage.
3. Under the null hypothesis, distributions for any two classes were identical, where two classes were randomly drawn from a larger observation space.
4. Under the alternative hypothesis, the probability that a sample from one class exceeding the probability of the other class was not 0.5.

The U test was applied similarly to the t test in which the observation of a candidate signal feature was separated into class 1 and class 0. We computed the U value of this signal feature and ranked it according to that value.

4.4.2 Computation

1. Combined $n1$ samples and $n0$ samples into one group W_i and order data in $W_1 \leq W_2 \leq .. \leq W_{n1+n0}$.
2. Denoted rank i for the i^{th} smallest observation.
3. Let R_1^{obs} be the sum of rank of observation in sample 1.
4. Calculated
$$K_1 = R_1^{obs} - \frac{n1(n1+1)}{2}. \tag{4.6}$$
5. Calculated
$$U_s^{obs} = max(K_1, n1n0 - K_1). \tag{4.7}$$
6. Calculated
$$U_s = max(R_1 - \frac{n1(n1+1)}{2}, n1n2 + \frac{n1(n1+1)}{2} - R_1) \tag{4.8}$$
where
$$\frac{n1(n1+1)}{2}$$
was the lowest value of R_1, and
$$n1n0 + \frac{n1(n1+1)}{2}$$
was the largest possible value.
7. Found the distribution of U_s under the null hypothesis, H_0. Reject if
$$P(U_s \geq U_s^{obs}) \leq \alpha,$$
where α was the one-sided p-value, by definition $U_s \geq \frac{n1n0}{2}$.

Individually, we computed the U value of all signal features with respect to this class of movement and repeated for the other classes. All signal features were ranked according their U value to select the top features.

4.5 Feature correlation criteria

In order to select next signal features after selecting the one with the highest ranking, we calculated feature correlation values. There were two values, namely the correlation information and regional information. It was applied to both t test and U test using z which represented t-value in the t test or U value in the U test.

The correlation information was calculated to find a dominant potential feature among all previously selected features

$$weights1 = z(1 - \alpha * rho) \tag{4.9}$$

where $\alpha = [0, 1]$ was the weighting factor in which an α close to 1 meant that signal features that are highly correlated to already selected signal features were less likely to be included, *rho* was an average absolute value of the cross-correlation coefficient between a candidate feature and a selected feature.

The regional information was calculated to find a dominant feature from the potential features based on the distance between a candidate feature and previously a selected feature. The weight was

$$weights2 = z(1 - \exp(-(\frac{d}{\beta})^2)) \tag{4.10}$$

where d was the L^2-norm distance between a candidate feature and a selected feature, β is the weighting factor, where β close to 0meant that features that were close to already chosen features were less likely to be included. The next feature was selected by

$$feature = \arg max_i(z_i * weights1 * weights2) \tag{4.11}$$

To adjust these parameters, a grid-search was employed. We examined 11 values of α in a range between 0 and 1 with a step of 0.1 and examined 11 values of β in a range between 0 and 2 with a step of 0.2. The objective function was the MLP which outputted the accuracy for evaluation. Evaluation was fulfilled by a 5-fold random cross-validation scheme which divided an entire set of data into five subsets in which one subset was used for training and the rest for testing. After all, we reported the average accuracy of five subsets.

4.6 Evaluating New Feature Sets by Pattern Recognition Algorithms

4.6.1 Input

The input for classification was the feature sets generated by the propose method. We used twenty sets of MES data (see chapter 3.2.10), each set was normalised by min-max range (0 to 1) and randomly divided into training, validation and test sets in the ratio of 70:15:15. We used the training set to obtain general structure of a classifier while the validation set was used to adjust tuning parameters of this classifier. The test set was separated from this process to evaluate the performance of the classifier using the achieved parameters in validation. Each dataset provided 11 classes of hand and wrist movements. A transient signal waveform corresponding to a class was segmented into 122 time segments. A time segment generated a feature vector using 21 TD features, an input vector of length 88 is described as shown in table 4.1.

Table 4.1: Elements of a feature vector.

Feature vector's element	From TD feature	At MES channel
1...4	tmn	1...4
5...8	tmabs	1...4
9...12	tmd	1...4
13...16	tstd	1...4
17...20	tvar	1...4
21...24	twl	1...4
25...28	trms	1...4
29...32	tzc	1...4
33...36	tpks	1...4
37...40	tmpks	1...4
44...48	tslpch1	1...4
49...52	tslpch2	1...4
53...56	tpwr	1...4
57...62	tcr	1...4
63...68	tcv	1...4
69...72	tdam	1...4
73...76	tfd	1...4
77...80	tmfl	1...4
81...84	tfdh	1...4
85...88	tren	1...4

4.6.2 Algorithms

In addition to the MLP used in chapter 3, we employed other pattern recognition algorithms, namely LDA and SVM. For all algorithms, the number of input nodes of the network matched the number of input elements in a feature vector (length of a feature vector). The number of output nodes of the network matched the number of required classified movements which were 11 classes mentioned in chapter 3.2.10. The performance metric was the classification rate or the accuracy of movement, i.e. the ratio of corrected classified movements to the total of required classified movements.

The MLP employed resilient backpropagation training. There was one hidden layer where the number of nodes matched the number of input nodes, the learning rate was 0.01, the weight change increment and decrement were 1.2 and 0.5, respectively. The activation function was tansig which outputted 1 for the target class and 0 for the others. All classes of movement were assigned at one time.

While the MLP was able to classify all classes at once, the LDA classified one class at a time (leave-one-out method). Similar to MLP, the target class was denoted by 1 and the other classes were denoted by 0.

The SVM employed a linear kernel function $k(x, y) = x.y$, and classification based on leave-one-out method. The optimisation method used was Sequential Minimal Optimisation (SMO). To adjust penalty parameter C, we applied two stages, a coarse stage and a finestage. Since the value of C was sensitive to the convergence of SVM training, the range of C was narrowed down in the coarse stage, in particular C received values 0.1, 0.3, 0.5, 0.7, 0.8, 0.9, 1, 1.1,1.2. In the fined stage, we examined C in the range of ±5%, ±10%, ±15% and ±20% of its selected value. We reported the accuracy in the experiments of the test set.

We used statistical test to report Wilcoxon Sign-Rank test p-values [18] on each pair of feature sets with a significant level of 5%. Performance of our feature sets was evaluated by the accuracy of hand and wrist movements, and compared to a typical feature set found in the literature which was: *tmabs, twl, tzc, tslpch2* [34, 59, 53]. The use of this feature set was recommended by Oskoei et al. [60] after a comparison among several single signal features and multiple feature sets in TD and FD. This feature set demonstrated the advantages over other feature sets in terms of a low computational load and stability over various time segment length.

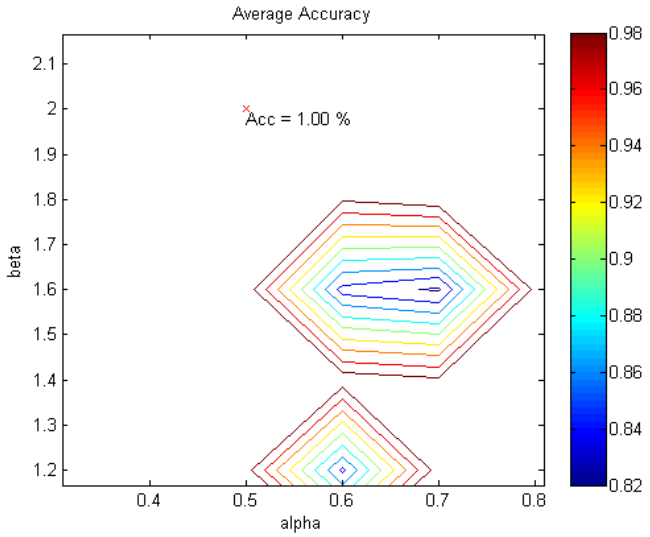
4.7 Results

4.7.1 Validation of tuning parameters for the search engines

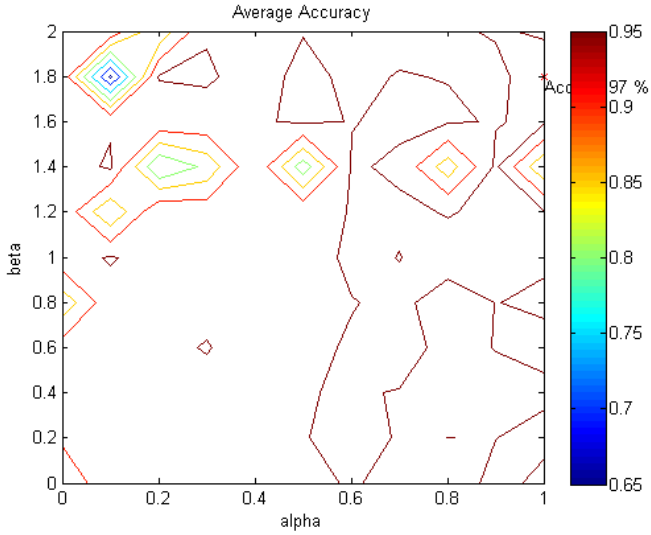
We validated α and β each time performing t test and U test to find the optimal feature correlation values. Fig. 4.1 shows an example of the grid-search to find the optimal pair of α and β. On the graph, the accuracy was represented by coloured line using the reddest colour for the highest accuracy and the bluest colour for the lowest accuracy. The colour bar on the right of the graph shows a range of accuracy from lowest to highest. An example illustrated on subject 20 shows that when $\alpha = 0.5$ and $\beta = 2$ the highest accuracy was 100% using t-test on MLP (Fig. 4.1a). In another example on subject 18, when $\alpha = 1$ and $\beta = 1.8$ the highest accuracy was 97% using U test on MLP (Fig. 4.1b).

4.7.2 Feature sets obtained by the three search engines

Fig. 4.2a shows the histogram of feature vector's elements selected by MI. As can be seen from the graph, there were three groups of dominant input element. The first group shows that elements 21,22,23,24 were highly selected corresponding to *wave length* (*twl*). The most frequent element was 22, counted for 17 times, and the lowest common element was 21, counted for 11 times. The second group of selection indicated that *maximum fractal length* (*tmfl*) dominated, corresponding to input elements 77, 78, 79, 80, cited by 9 times and 16 times of selection for the lowest and the highest frequent elements, respectively. The third group contained input elements 85,86,87,88 which occupied equal frequency of selection at 19 times. This

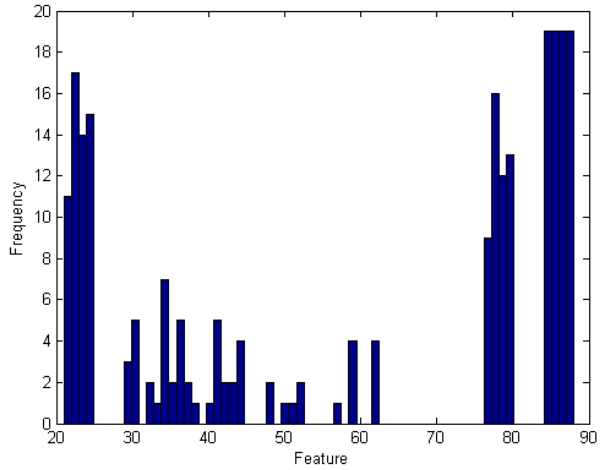


(a) Two-sample t-test

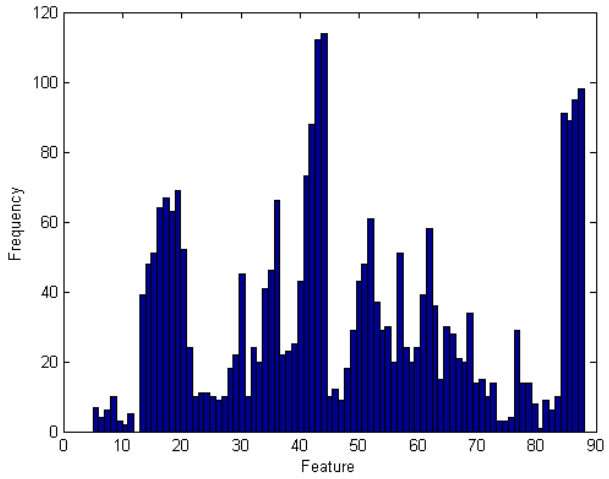


(b) Mann-Whiteney U test

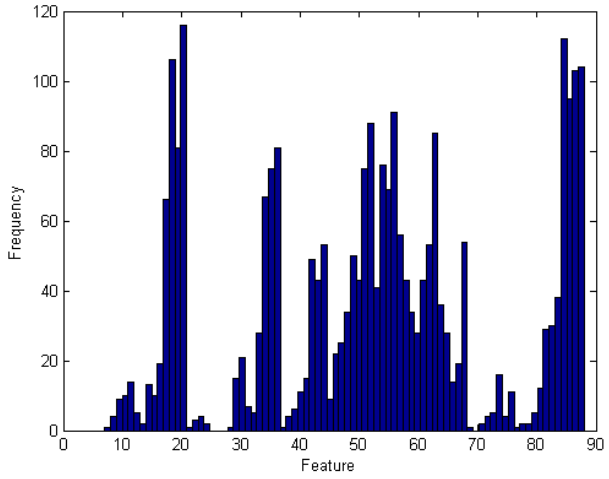
Figure 4.1: Grid-search of α and β on t test (a) and U test (b) performed by cross-validation accuracy. This contour plot indicates the position of the highest accuracy (marked by 'x') corresponding to value of α on the x-axis and value of β on the y-axis.



(a) Mutual information



(b) Two-sample t-test



(c) MannWhitney U test

Figure 4.2: Histogram of feature selected by MI, Two-sample t-test, and MannWhiteney U test.

group represented the *rough entropy per channel* (*tren*).

In short, MI selected three TD features which were *wave length* (*twl*), *maximum fractal length* (*tmfl*), and *rough entropy per channel* (*tren*). We used these signal features as a feature set for classification, named as *set 1*.

Fig. 4.2b shows the histogram of feature vector's element chosen by the t test. The graph reveals three groups of element. Group 1 contained input elements of two TD features, element 16 belonged to *standard deviation* (*tstd*), and 17,18,19 belonged to the *variance values* (*tvar*). They were cited by more than 60 times of selection. Group 2 consisted of elements 41,42,43,44 corresponding to the *mean of the differences of MES peak voltage over RMS value* (*tmvel*), cited by over 70 times of selection. Group 3 involved elements 85,86,87,88 representing the *rough entropy per channel* (*tren*), repeated by over 80 times of selection.

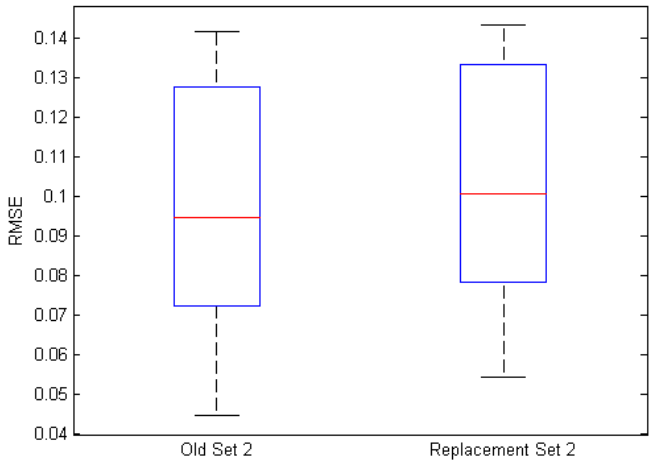
In summary, the t test accepted three TD features for a feature set, namely *standard deviation* (*tstd*), *mean of the differences of MES peak voltage over RMS value* (*tmvel*), and the *rough entropy per channel* (*tren*). This feature set was referred as *set 2*.

Fig. 4.2c shows the histogram of feature vector's element selected by the U test. The graph depicts high frequency of several input elements which did not completely represent four MES channels of a signal feature. For simple extraction, we assumed the use of the entire four MES channels of a signal feature. The selection of signal feature was based on this assumption. There were four groups of input elements. The first group of elements, cited by high frequency of selection, from 65 times to 110 times, contained elements 17,18,19,20. They represented the *variance of MES data* (*tvar*). The second group comprised three elements 34,35,36 which were the *number of MES peak voltage over MRS value* (*tpks*). This group had similar frequency of selection from 70 times to 80 times. The third group included three input elements 54,55,56 corresponding to the *power of MES data* (*tpwr*), counted by about 70 times of selection. The fourth group contained elements 85,86,87,88 corresponding to *rough entropy per channel* (*tren*), cited by over 100 times of selection.

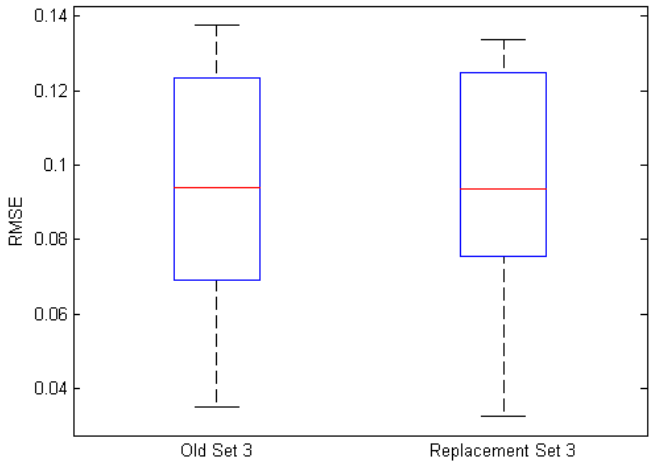
In brief, the results of U test suggested four TD features for a feature set including the *variance of MES data* (*tvar*), *number of MES peak voltage over MRS value* (*tpks*), the *power of MES data* (*tpwr*), and *rough entropy per channel* (*tren*). We named this feature set as *set 3*.

It is noticed that a signal feature generates four input elements corressponding to four MES channels. However, the experiments showed that selection based on three class relevant criteria did not recommend a complete four channel of a signal feature, indicated by the histogram in Fig. 4.2. Therefore, we made an assumption that if a signal feature was selected, a complete four elements belong to that signal feature would be selected. This asummption was applied to assemble the above feature sets.

To test our assumption, we examined the root mean square errors (RMSEs) of the actual features recommended by t test and U test, named as Old set 2 and Old set 3, respectively; and compared to the RMSEs of the assumed feature sets, known as Set 2 and Set 3 mentioned above. Fig. 4.3 shows that there was no significant difference of RMSEs between the Old set (on the left of the graph) and the assumed set (on the right of the graph).



(a) Old set 2 and replacement set



(b) Old set 3 and replacement set

Figure 4.3: Boxplot of set 2 and its replacement set (a), set 3 and its replacement set (b). The central mark is the median, the edges of the box are 25th and 75th percentiles, the whikers are extreme data points where ends of the whiskers are minima and maxima.

Table 4.2: Results of set 1 and a literature set on average accuracy of 11 movements on 20 subjects.

	Set 1 [a]		Literature set [b]		
Algorithm	Computational time (s)	acc (%) ±SD	Computational time (s)	acc (%) ±SD	Wilcoxon p-value
MLP	0.380 ±0.204	93.96 ±0.034	0.332 ±0.097	90.06 ±0.060	8.8575e-05
LDA	0.176 ±0.035	91.48 ±0.048	0.205 ±0.014	90.03 ±0.053	0.0365
SVM	0.128 ±0.032	95.78 ±0.021	0.233 ±0.015	95.42 ±0.024	0.0522

[a] *twl, tmfl, tren*
[b] *tmabs, twl, tzc, tslpch2*

Table 4.3: Results of set 2 and a literature set on average accuracy of 11 movements on 20 subjects.

	Set 2 [a]		Literature set [b]		
Algorithm	Computational time (s)	acc (%) ±SD	Computational time (s)	acc (%) ±SD	Wilcoxon p-value
MLP	0.942 ±0.104	93.20 ±0.039	0.332 ±0.097	90.06 ±0.060	0.0013
LDA	0.324 ±0.017	92.47 ±0.023	0.205 ±0.014	90.03 ±0.053	0.0072
SVM	0.305 ±0.028	96.16 ±0.022	0.233 ±0.015	95.42 ±0.024	0.1005

[a] *tvar, tmvel, tren*
[b] *tmabs, twl, tzc, tslpch2*

Table 4.4: Results of set 3 and a literature set on average accuracy of 11 movements on 20 subjects.

	Set 3 [a]		Literature set [b]		
Algorithm	Computational time (s)	acc (%) ±SD	Computational time (s)	acc (%) ±SD	Wilcoxon p-value
MLP	0.820 ±0.103	94.07 ±0.037	0.332 ±0.097	90.06 ±0.060	1.2042e-04
LDA	0.267 ±0.086	92.74 ±0.023	0.205 ±0.014	90.03 ±0.053	0.0040
SVM	0.235 ±0.015	96.26 ±0.019	0.233 ±0.015	95.42 ±0.024	0.0072

[a] *tvar, tpks, tpwr, tren*
[b] *tmabs, twl, tzc, tslpch2*

In conclusion, we generated three feature sets from three class relevant criteria. They were set 1 = {*twl, tmfl, tren*}, selected by MI; set 2 = {*tstd, tmvel, tren*}, selected by t test; and set 3 = {*tvar, tpks, tpwr, tren*}, selected by U test.

4.7.3 Performance of the new sets of features on movement classification

Performance metric for classification was the accuracy. Table 4.2 compares the average accuracy of set 1 and a typical literature set of features. The accuracy was calculated over 20 subjects to classify 11 hand and wrist movements. A remarkable result was that set 1 outperformed the literature set on three algorithms MLP, LDA and SVM. The improvement had a beneficial effect on MLP and LDA, as demonstrated by $p < 0.05$. On SVM, however, either set 1 or the literature set was favourable.

Table 4.3 presents a comparison of accuracy between set 2 and the literature set. It

was similar to that of set 1 when set 2 was superior. In addition, signifiant small p values shows that both MLP and LDA benefited from this set while SVM did not.

On the other hand, set 3 outferformed the literature set over three pattern recognition algorithms, as shown by minimal p values in Table 4.4. Moreover, set 3 achieved higher accuracy than both set 1 and set 2. This can be explained by the higher input elements it provided to the classifiers, given by four inputs instead of three for those in set 1 and set 2.

Time taken to achieve these accuracies in three sets were similar to that of the literature set which was less than 0.4 seconds when performed by LDA and SVM. For MLP, however, computational times of set 2 and set 3 were longer at 0.942 second and 0.820 second, respectively, compared to 0.332 second for the literature set.

To sum up, performance details of three acquired feature sets was shown in Fig. 4.4, indicating the accuracy measured on individual subject and individual movement. These feature sets were effectively superior to the literature set throughout the subjects. They significantly improved the accuracy of subject 13 and 14 where the performance of the literature set was depressed. Although a narrow accuracy of some movements such as side grip (SG), fine grip (FG), agree (AG), pointing (PT); the acquired feature sets did rise the accuracy of these movements. In particular, the accuracy rose from under 80% of the literature set to 87% of set 1, as shown for example of SG on MLP.

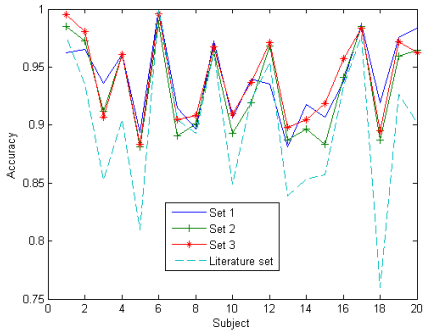
In conclusion, our proposed methods generated three feature sets that successfully increased the accuracy of 11 hand and wrist movements. They can be used as alternative solutions for FS using TD features.

4.8 Conclusion

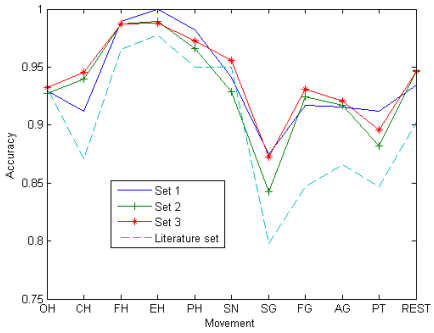
In this chapter, we propose a FS method using class relevant coefficient criteria. Three criteria namely MI, t test, and U test are consecutively employed to assemble three feature sets from twenty one potential TD features. The acquired feature sets are superior to a typical feature set found in the literature, demonstrating the possibility of alternative feature sets in the time domain. The advantages of these sets are shown on MLP and LDA as these classifiers consistently accomplish higher accuracy than that of the literature set over subject and movement.

In addition, there are two advantages of this method compared to GA. First of all, the feature sets resulting from this method are comprehensive in terms of length. They only require three to four signal features compared to more than ten features of GA. Secondly, these feature sets can be used for several subjects, and are suitable for many pattern recognition algorithms.

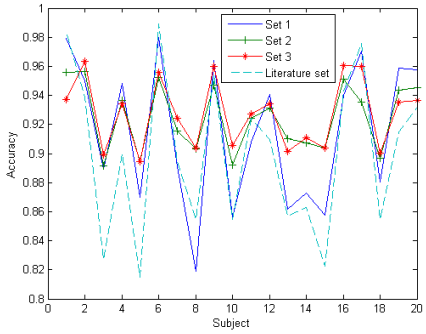
A major contribution of this chapter is that our proposed method demonstrates that the FS approach is not limited to meta-heuristic search. This is the first time such statistical tests are applied in FS by investigating the response of signal features to movement classes, and that makes a contribution to broaden the data preprocessing horizons of prosthetic control.



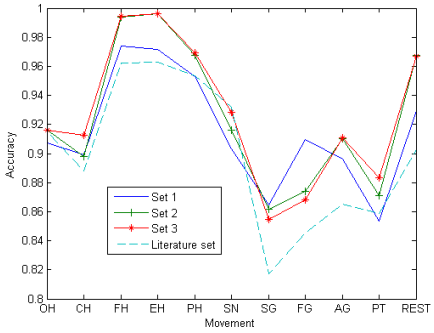
(a) MLP on subject



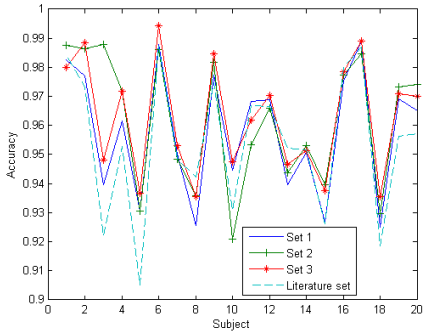
(b) MLP on movement



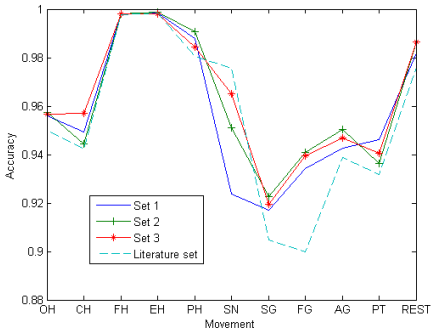
(c) LDA on subject



(d) LDA on movement



(e) SVM on subject



(f) SVM on movement

Figure 4.4: Accuracy of four feature sets measured on subject and movement. The experiement employed MLP, LDA and SVM.

Chapter 5

Self Organising Feature Map and Feature Contribution

5.1 Introduction

In the previous chapters, TD features were compared to the others within each time segment for pattern structure preservation. This, however, prevents us from measuring intra-class feature stability. In this chapter, we aim to select relevant signal features using a Self Organising Feature Map (SOFM) along the MES transient waveform. Twenty one TD features previously segmented are joined to time series vectors. They are mapped to clusters by the SOFM. The network is implemented by optimising network dimension, weight vectors initialisation and learning algorithm. They are summarised as follows.

We calculate the Davis-Bouldin (DB) index to find the maximum number of clusters demanded for mapping twenty one TD features. The numbers of clusters indicates the number of neurons, thus determine the network dimension.

The initialisation of weight vectors is useful to accelerate learning speed. Because of a high dimension of the input vector which consists of twenty one TD features, it challenges the SOFM. This problem can be overcome by arranging the weight vectors in a reasonable order following the input. Implementation is taken by the singular value decomposition of the input difference from the mean (IDM).

The learning algorithm employs batch learning in which weight vectors are updated at the end of an epoch when all inputs are fed in the network. This learning strategy is particularly useful in our implementation programming language which is MATLAB R2013b.

After being mapped by SOFM, signal features in each cluster are ranked according to our implemented index of contribution percentage (CP). The highest ranked feature in a cluster is selected. We assemble a large feature set from these highest ranked features, and calculate CP for the second time. A smaller size of feature set is chosen so that it remains at least 95% accuracy of the entire set.

Three pattern recognition algorithms, namely MLP, LDA, and SVM are employed

Table 5.1: Four input vectors of a signal feature f_i

Signal feature f_i			
Channel 0	Channel 1	Channel 2	Channel 3
$x_{i0,1}$	$x_{i1,1}$	$x_{i2,1}$	$x_{i3,1}$
...	...	...	...
$x_{i0,122}$	$x_{i1,122}$	$x_{i2,122}$	$x_{i3,122}$

to evaluate the performance of the above small feature set, and compare it with a typical feature set found in the literature. We report the Wilcoxon p value of the statistical test to substantiate our experimental results. A feature set drawn from this method, referred as set 4, improves the accuracy of hand and wrist movements.

This chapter is organised as follows. Section 5.2 explains the implementation of the SOFM. Section 5.3 lists three types of input vector mapped by the SOFM. Section 5.4 provides the calculation of the CP of a signal feature. Section 5.5 briefly states three classifiers used for classification which was defined previously in chapter 4. Section 5.6 reports the experimental results. Section 5.7 concludes the chapter.

5.2 Self Organising Feature Map

We considered a single non-linear SOFM in which its parameters were represented by

$$y = \hat{f}(x, w) \tag{5.1}$$

where x was the input, and y was the output; $x = (x_1, ..., x_R)^T$ was the time series vector extracted from a signal feature divided into R segments, w was the weight vector. Given N neurons, neuron i was determined by $w_i = (w_{i1}, ..., w_{iR})^T$, with $i = \{1, ..., N\}$.

There were 122 time segments, $R = 122$, corresponding to the length of an input vector. A signal feature fi produced four input vectors corresponding to four MES channels. They are denoted in table 5.1.

5.2.1 Evaluation

To draw a feature set from twenty one signal features, we evaluated individual signal features. Given m signal features, the evaluation followed these steps:

1. Mapped time series vector $x_{i,n}$ of feature f_i to set M_i using SOFM

$$M_i = \{x_{i,n} : i = \{1, ..., m\}, n = \{0, 1, 2, 3\}\}$$

2. Calculated RMSEs by running the MLP for set M_i.

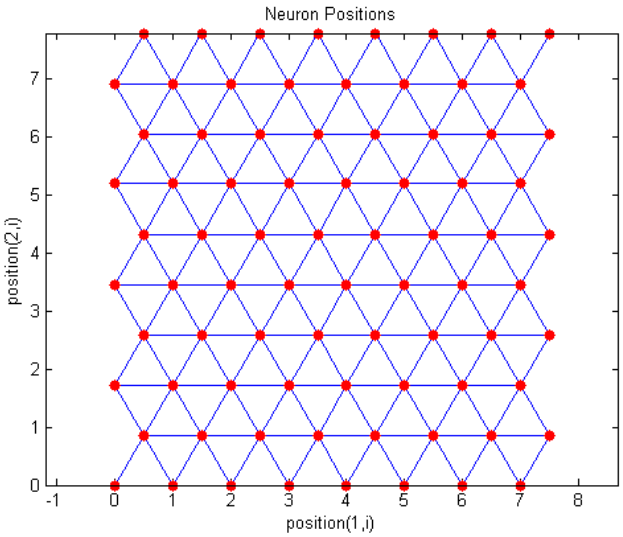


Figure 5.1: Initial an 8-by-10 set of neurons in a hexagonal topology.

3. Calculated contribution percentage (CP) of $x_{i,n}$ in each M_i.
4. Ranked $x_{i,n}$ according to its CP index.
5. Selected the highest ranked $x_{i,n}$ and formed a new set.
6. Went back to Step 2 and repeated the procedure for the second time on the new set.

We drew a feature set based on the final ranked set. An optimal feature set was a small set that was expected to retain at 95% of the accuracy of the full set.

In the next sections, we will describe SOFM, CP and report experimental results.

5.2.2 Topology of SOFM

The original neuron locations were organised by *hextop* function.The neurons were arranged in a hexagonal layer geometry (Fig. 5.1).

The distance between neurons was determined by their links, i.e. the number of connections or steps needed to get a neuron under the other's neighbourhood size. Given two position vectors $p_i = (p_{i1}, ..., p_{in})^T$ and $p_j = (p_{j1}, ..., p_{jn})^T$ from a set of S vectors, link distance was computed by

$$
\begin{aligned}
d_{ij} &= 0, \quad \text{if } i = j \\
&= 1, \quad \text{if } \sqrt{\sum_{l=1}^{n}(p_{il} - p_{jl})^2} <= 1 \\
&= 2, \quad \text{if } k \text{ exists, } d_{ik1} = d_{k2j} = 1 \\
&= 3, \quad \text{if } k1, k2 \text{ exist, } d_{ik1} = d_{k1k2} = d_{k2j} = 1 \\
&= N, \quad \text{if } k1, .., kN \text{ exist, } d_{ik1} = d_{k1k2} = ... = d_{kNj} = 1 \\
&= S, \quad \text{if none of the above conditions applied}
\end{aligned}
\tag{5.2}
$$

where $k = 1, .., S$ denoted position vector $\{P_k = (p_{k1}, ..., p_{kn})^T | \, d_{ik} = d_{kj} = 1\}$

5.2.3 Network dimension

The initial network dimensions started with 10 neurons, in an order of a two dimensional 2-by-5 lattice, and was increased by a step of 5 neurons until redundant neurons appeared. To recognise unnecessary neurons, we computed Davis-Bouldin (DB) index [16]

$$DB = \frac{1}{N}\sum_{i=1}^{N}\max_{j\neq i} D_{ij}. \tag{5.3}$$

The optimal number of clusters would minimise this average similarity. D_{ij} indicated the similarity of two clusters, and N was the number of clusters. The bounded condition DB for a maximum number of clusters was calculated by a maximum value of D_{ij}, where

$$D_{ij} = \frac{S_i + S_j}{d_{ij}} \tag{5.4}$$

where S_i, S_j were the average Euclidean distances within clusters. They were computed by

$$S_i = \frac{1}{N_i}\sum_{n=1}^{N_i} ||\overline{x_i} - x_{i,n}|| \tag{5.5}$$

$$S_j = \frac{1}{N_j}\sum_{n=1}^{N_j} ||\overline{x_j} - x_{j,n}|| \tag{5.6}$$

where N_i and N_j Nj were the number of samples $x_{i,n}$, $x_{j,n}$ mapped to cluster i and j, respectively, $\overline{x_{(.)}}$ was the centre of cluster which was the mean of $x_{(.),n}$. The distance between clusters was

$$d_{ij} = ||\overline{x_i} - \overline{x_j}||. \tag{5.7}$$

The Davis-Bouldin index showed the maximum number of clusters or neurons required to map the input. Hence, the network dimension of SOFM was

$$\text{network dimension} = \{\min DB \cap \min \textit{empty neuron}\}. \tag{5.8}$$

5.2.4 Network initialisation

The network initialisation comprised the initial values of the network dimensions, original neighbourhood size nd_0, ordering step τ_1 and the neuron's normalisation. While the other parameters would be adjusted during experiment, the neuron's normalisation was initialised in the position normalisation which initialised weight vectors.

Position normalisation

This step initialised weight vector positions with: the weight vector distributed across the input space. The SOFM learning could be accelerated by implementation of initial position of weight vectors. We set up the weight vectors commencing at a acceptable distribution over the input mapping. The function employed was *initsompc*. The weight vectors were influenced by the inputs, the map dimensions and initial neuron positions as described below

$$weight = f(inputs, dimensions, positions). \tag{5.9}$$

Since the inputs and map dimension were the outlier factors, we adjusted the initial neuron position. Mathematically, the weight spread was

$$weight = posMean + posBasis * pos. \tag{5.10}$$

Each variable was computed as follows. Firstly, we calculated *posMean*.

1. *posMean*
   It was an R-by-1 matrix of the mean value of input's samples, where R was the number of input's elements, R = 122. Given Q samples,

   $$posMean = (x_{1,n}, ..., x_{i,n}, ..., x_{R,n})^T \tag{5.11}$$

   where

   $$x_{i,n} = \frac{1}{Q} \sum_{n=1}^{Q} sample_n.$$

2. *posBasis*
   It was an R-by-R matrix of the basis of input vector's position. Firstly, we calculated the input difference from mean (IDM)

$$IDM = x - posMean. \tag{5.12}$$

Secondly, we found U, Σ, V which were the components of the *singular value decomposition* of IDM such that $IDM = U\Sigma V^T$; and V^T was the transpose of V.

Thirdly, the basis

$$basis = U * \Sigma \tag{5.13}$$

Next, the standard deviation of V

$$std = \left(\frac{1}{Q} \sum_{i=1}^{Q} (V(i) - \overline{V})^2 \right)^{\frac{1}{2}} \tag{5.14}$$

where $\overline{V}$ was the mean value of V.

Finally,

$$posBasis = \sum_{i=1}^{R} \sum_{p=1}^{R} basis(p, i) * 2.5 * std(i). \tag{5.15}$$

3. *pos*
   It was the normalised input's position. The positions were normalised by mapping min-max position values to the $[-1, +1]$ interval

$$pos = 2\frac{pos - minPos}{maxPos - minPos} - 1 \tag{5.16}$$

   where *minPos*, *maxPos* were minimum and maximum position values, respectively.

When all variables were determined, the initial position of weights was calculated by Eq. 5.10.

5.2.5 Learning Procedure

The SOFM learning procedure consisted of 3 steps: competitive, cooperative and adaptive process. This procedure was constructed by using a batch learning algorithm. The basic concept was that in each training cycle, the weights were updated at the end of the epoch. In particular, all inputs were presented for training in an epoch, and all the weights were adjusted according to the average of samples fed into the network. This algorithm worked effectively in the MATLAB R2013b which ensured quick convergence. The learning function was *learnsomb* which was a MATLAB Neural Network Toolbox version 8.0 64-bit. Below is a brief description of each process.

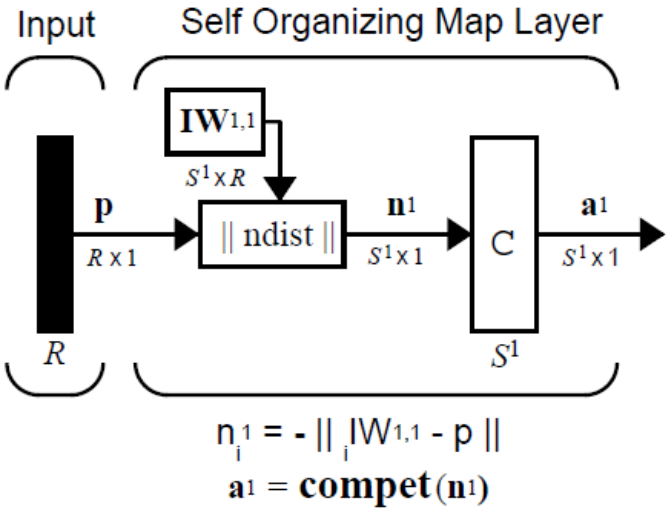


Figure 5.2: SOFM one layer architecture [50]

Competitive process

A competitive process determined the winning neuron. The SOFM was constructed by a single neuron layer architecture including an input layer and an output layer which was the neuron layer. The SOFM had no hidden layer (Fig. 5.2). The network included no bias. In the competitive phase, a transfer function produced 1 corresponding to the winning neuron i and for all the neurons close to it. For example, in 5.2, when the input vector p was presented, the network determined a winning neuron n_i^1 (layer1) by computing the distance using $||ndist||$. The transfer function denoted by *compet* outputted a $a_i^1 = 1$ corresponding to n_i^1 and for all surrounding neurons within a deterministic neighbourhood distance. The rest of the outputs were 0. The distance function $||ndist||$ was set to the negative Euclidean distance (*negdist*) which was

$$z = -\sqrt{\sum_{l=1}^{S}(w-x)^2}. \tag{5.17}$$

The SOFM defined the best matching unit (winning neuron), denoted by b, was the mapped unit which was closest to the input x

$$||w_b - x|| = \arg\min_i ||w_i - x|| \tag{5.18}$$

where w_b was the weight vector of the best-matching unit (BMU). The output was represented by $y = (\{y_i : i = 1, ..., N\})^T$ where $y_i = 1$ when $i = b$ and $y_i = 0$ when $i \neq b$, and $||.||$ was the l_2 norm or the Euclidean distance. In this application, $||ndist|| = negdist$, thus the BMU was determined by

$$\|w_b - x\| = \arg\max_i\{-\|w_i - x\|\}. \tag{5.19}$$

Cooperative process

This process determined all neurons located within the neighbourhood distance nd to be activated and be updated their weights. The *nd* startted with a given value and reduced during the *ordering phase* of the *adaptive process*. The entire process could be divided into two steps:

- Step 1: Found neighbour neurons of the BMU
  To find these neurons, we computed the neighbourhood distance *nd*, it shrank by a function of time

$$nd(n) = nd_{final} + (nd_0 - nd_{final})(1 - \frac{n}{\tau_1}) \tag{5.20}$$

  where nd_0 was the original neighbourhood distance (default was $nd_0 = 3$), τ_1 determined the number of steps that the *ordering phase* would last for (default was$\tau_1 = 100$). The computation of *nd* would repeat in the *ordering phase* where *nd* would reduce to a deterministic value before the network turned into the *tuning phase*, typically at $nd_{final} = 1$ so that only the winning neuron was updated. The neighbourhood distance became

$$nd(n) = 1 + (nd_0 - 1)(1 - \frac{n}{\tau_1}). \tag{5.21}$$

  The neighbours of the BMU satisfied the condition

$$\begin{aligned} neigbourhood = n_{ij} = 1 \quad & \text{if } d_{ij} <= \sigma_n, \\ = n_{ij} = 0 \quad & \text{if } d_{ij} > \sigma_n. \end{aligned} \tag{5.22}$$

  where d_{ij} was the distance between neurons i and j, the neurons within the neighbourhood were denoted by 1, the others by 0.

- Step 2: Activated neurons within the neighbourhood
  A random 90% of the previous activated output was select

$$a(n) = a(n-1) * rand(a(n-1) < 0.9) \tag{5.23}$$

  where the output vector a was a logical N-by-Q matrix, N was the number of neurons, Q was the number of input's samples, in this case Q = 88, $rand(a(n - 1) < 0.9)$ was a random function selecting 90% elements of a. The activated neurons were

$$a2 = neighbourhood * a(n) + a(n). \tag{5.24}$$

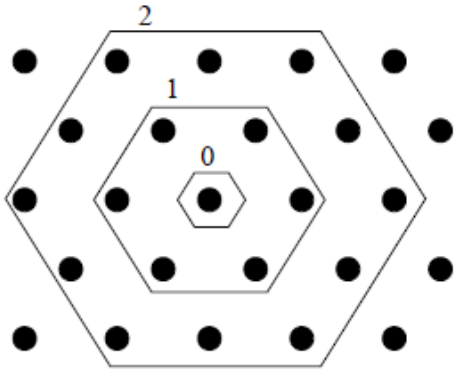


Figure 5.3: Neighbourhood size (size 0, 1, 2) of the central neuron. The smallest polygon corresponds to size 0 - no neighbourhood, the second to size 1, and the biggest to size 2.

Each row of $a2$ represented a neuron while each column represented a sample. The values of $a2$ were

$$\begin{aligned} a2(i,q) &= 2, && \text{if } a(i,q) = 1 \\ &= 1, && \text{if } a(j,q) = 0 \text{ and } d(i,j) <= nd \\ &= 0, && \text{otherwise.} \end{aligned} \tag{5.25}$$

In other words, $a2(i,q) = 2$ corresponds to the winning neuron, $a2(j,q) = 0$ represents neighbourhood neurons, and $a2(i,q) = 0$ represents the rest.

Adaptive process

Here we updated the weights according to Kohonen's learning rule for batch algorithms [42]. It was required that the expected value of $w_i(n)$ and $w_i(n-1)$ must be equal when $n \to \infty$. Therefore, $w_i(n)$ and $w_i(n-1)$ could be replaced by w_i^*.

$$w_i^* = \frac{\sum_{q=1}^{Q} a2(i,q) * x^T}{\sum_{q=1}^{Q} a2(i,q)} \tag{5.26}$$

note that

$$\begin{aligned} w_i^* &= w_i^*, && \text{if } \sum_{q=1}^{Q} a2(i,q) \neq 0 \\ &= 0, && \text{otherwise} \end{aligned}$$

where x^T was transposed matrix of x, an R-by-Q input matrix in which R was the number of segments, Q was the number of input's samples; $R = 122$, $Q = 88$.

The network was trained by *trainbu*, the batch unsupervised learning function. There was no bias in SOFM learning but only the weights. A weight was updated by Eq. 5.26.

The training was terminated when one of these parameters exceeded: time frame, the number of epochs, performance minimised to the goal, or the number of validation times reaching the maximum value since the last time it decreased.

The adaptive process also defined the neurons mapping. First, it generalised all neurons according to the input space. Second, it refined the map by tuning only winning neurons. Consequently, adaptive process was divided into an *ordering phase* and a *tuning phase*.

- Ordering phase
  In this phase, the neighbourhood function *nd* started at a given initial distance nd_0, and decreases to 1 (nd_{final}) followed Eq. 5.21. This phase lasted for a number of steps, specified by τ_1. During the time that the neighbourhood function reduced, the neurons of the network were expected to order themselves in the input space with the same topology as the input patterns.

- Tuning phase
  This phase lasted for the rest of the training. There was no learning rate required in batch algorithm, only the neighbourhood function *h*. In addition, *h* was cancelled in the original weight updating equation, as proved in [42], and replaced by Eq. 5.26. It is noticed that *h* was existing and it had decreased to 1 in the former phase. Therefore, a winning neuron adjusted itself while the neighbourhood neurons maintained the previous arrangement. At this phase, the network was expected to be considerably well ordered. The training continued so that the neurons had time to distribute steadily over the input space.

5.3 Mapping features

In general, the SOFM mapped a time series feature vector to a group. As mentioned at the beginning of section 5.2, a signal feature generated four input vectors. These vectors might not be mapped to the same group. In other words, they might all have similar characteristics or might share its characteristics with other signal features. The SOFM identified three types of sharing information (Fig. 5.4).

- Type 1: all time series vectors of the feature had the same and unique characteristic

$$\{x_{i,n} : n = 0, 1, 2, 3\} \subseteq f_i.$$

- Type 2: at least one time series vectors of feature f_i overlapped the time series vectors of feature f_j

$$\{x_{i,n} : n = 0 \vee 1 \vee 2 \vee 3\} \subseteq (f_i \cap f_j).$$

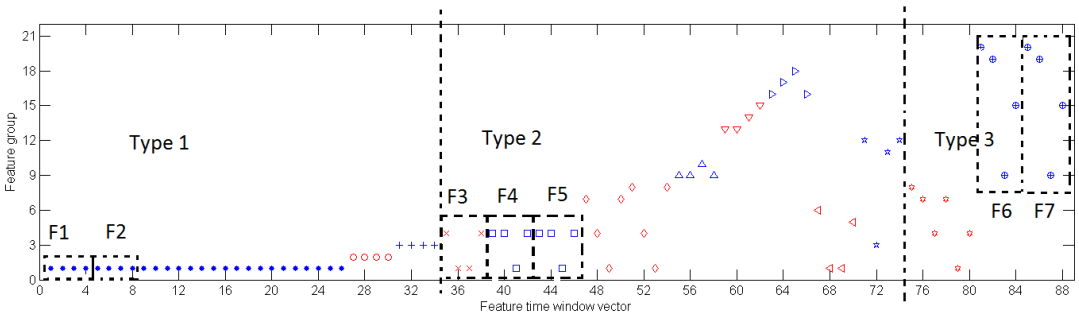


Figure 5.4: Three types of signal features identified by single-layer SOFM. A marker represents a time series vector. The same level of markers indicates similar vectors. For example, feature F1 and F2 are identical; F4 and F5 are identical and in intersection of F3; F6 and F7 are identical and being subset of other features.

- Type 3: all time series vectors of feature i overlapped the vectors of other features

$$\{x_{i,n} : n = 0, 1, 2, 3\} \subseteq (f_i \cap f_j), i \neq j.$$

In a situation that two signal features with similar CP values were mapped to the same group, based on different types of sharing information, we decided which signal feature were more important. Hence, when CP values were similar:

1. If two features in type 1, they could replace each other.
2. If two features in type 2, they were reserved.
3. If two features in type 3, they could replace each other.

The calculation of CP value was described in the following section.

5.4 Feature Contribution

The contribution percentage (CP) of an individual signal feature was evaluated by root mean square errors (RMSEs) instead of the accuracy. It was because an accuracy level could be specified by a wide range of RMSEs. A signal feature was ranked according to its CP in descending order. The CP was

$$CP = \frac{\Delta err_{rmsi}}{\sum_{i=1}^{n} \Delta err_{rms}} \times 100 \tag{5.27}$$

where

$$\Delta err_{rmsi} = err_{rmsi} - err_{rmsOrg}$$

$$\Delta err_{rmsi} = \begin{cases} \Delta err_{rmsi} & \textbf{if } \Delta err_{rmsi} \geq 0 \\ 0 & \textbf{if } \Delta err_{rmsi} < 0 \end{cases}$$

where err_{rmsi} was the RMSEs associated with feature f_i omitted, err_{rmsOrg} was the original RMSEs without feature omission, n was was the number of signal features in a group.

5.5 Evaluation of feature set on movement classification

The feature set found by this method was evaluated by pattern recognition algorithms. The performance metric was the accuracy of movements. The classifiers used were MLP, LDA and SVM. See chapter 4.6.2 for a description of these algorithms. The found feature set was compared to a typical literature set as mentioned in chapter 4.6.2.

5.6 Results

5.6.1 Network initialisation

We set up the initial neighbourhood size at $nd0 = 3$ and the ordering step $\tau_1 = 100$. The original position of neurons was calculated and distributed using Eq. 5.10. A below paragraph is the initialisation of network dimensions.

Network dimension

The network dimensions satisfied condition 5.8. We started with 10 neurons by a two dimensional lattice of 2-by-5 and increased until one of the terminal conditions met. Fig. 5.5 shows the DB index of the SOFM network (top) and the number of empty neurons (bottom). As can be seen from the graph, the DB index dropped steadily when the number of neurons increased. An empty neuron occurred at 28. For ease of computation, we selected a network of 25 neurons which was a lattice of 5-by-5.

5.6.2 Mapping features

Table. 5.2 shows signal features mapped by the SOFM. As expected, type 1 was a collection of identical features. There were four groups of signal features in this type. Group 1 reflected signal amplitudes including the *mean and median of data* (*tmn* and *tmd*). Group 2 was the probability distribution and frequency sensitivity of the MES data containing the *variance* (*tvar*), *slope change over RMS* (*tslpch2*), *power* (*tpwr*) and *mean of the difference of data* (*tdam*). In group 3, *Katz's fractal dimension* (*tfd*) was mapped by itself and group 4 was *Higuchi's fractal dimension* (*tfdh*).

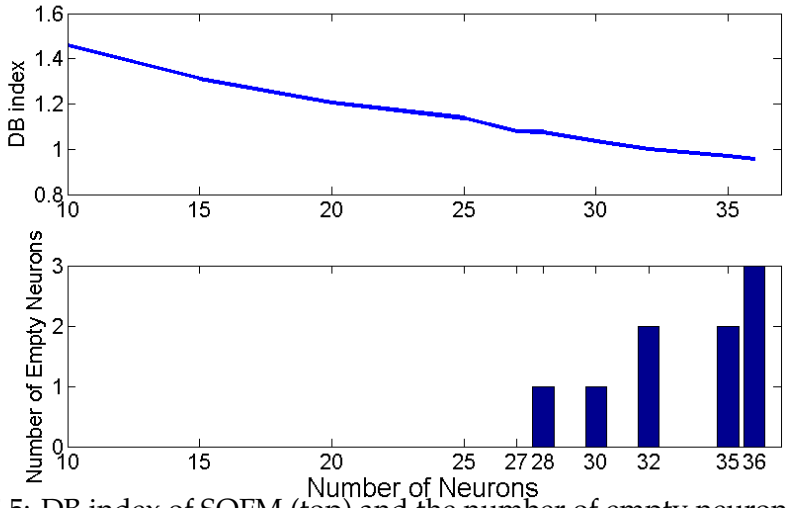


Figure 5.5: DB index of SOFM (top) and the number of empty neurons (bottom)

Type 2 comprised partly shared information features including seven groups. Some groups contained signal features which shared the information within group such as group 1,2,3,4. The other groups embraced features sharing information between group such as group 5,6,7. Type 3 only consisted of one signal feature which was the *zero crossing* (*tzc*). It comprised four input vectors overlapped the input vectors of four other signal features.

5.6.3 Feature contribution and evaluation

Table 5.3 shows CP value and group ranking of signal features previously mapped. We selected the highest ranked within group which were *tmd, tdam, tmfl, tmpks, tstd*. A final feature set was drawn from these signal features and eight individuals that were previously mapped in single groups. The final feature set comprised *tmd, tdam, tmfl, tmpks, tstd, tfd, tfdh, tpks, tslpch1, tren, tzc, twl, tmvel*.

Table 5.4 depicts the ultimate ranking and CP value of signal features in the final feature set. Top three highest ranked was *rough entropy per channel* (*tren*), *correlation of MES data* (*tcr*), and *zero crossing* (*tzc*).

Lastly, we reported a range of accuracy of several narrow feature sets drawn from the final feature set. Fig. 5.6 shows the accuracy range from a feature set with one omitted signal feature to a set with twelve omitted signal features. The test was done by MLP, trained by resilient backpropagation (*trainrp*) and Mevenberg-Marquardt back-propagation (*trainlm*).

For *trainrp*, the final feature set of full thirteen signal features achieved the accuracy of 97.25% and it fell to 95.54% with a feature set of three comprising *tren, tcr,* and *tzc*. For *trainlm*, these three signal features accomplished the accuracy of 97.08% which retained very high accuracy compared to the final feature set. As a result, we accepted this feature set, named as *set 4* containing *tzc, tcr tren*.

Table 5.2: Signal features mapped by SOFM.

	Type 1	Type 2	Type 3
Group 1	tmn,tmd	twl,tmfl	tzc
Group 2	tvar, tslpch2, tpwr, tdam	tmpks,tmvel	
Group 3	tfd	tmabs, tstd, trms	
Group 4	tfdh	tcr,tcv	
Group 5		tpks	
Group 6		tslpch1	
Group 7		tren	

Table 5.3: Ranked features in each group according to CP index.

Ranking	Group	RMSE ±SD	CP (%)
1	tmd	0.287 $\pm 7\times10^{-4}$	83.01
2	tmn	0.268 ±0.008	16.99
1	tdam	0.216 ±0.019	76.29
2	tvar	0.176 ±0.023	13.39
3	tpwr	0.174 ±0.027	10.26
4	tslpch2	0.156 ±0.028	0.05
1	tmfl	0.201 ±0.016	52.05
2	twl	0.198 ±0.018	47.95
1	tmpks	0.199 ±0.023	52.17
2	tmvel	0.198 ±0.017	47.83
1	tstd	0.165 ±0.025	47.33
2	trms	0.159 ±0.022	27.75
3	tmabs	0.157 ±0.024	24.70
1	tcr	0.254 ±0.009	66.24
2	tcv	0.223 ±0.018	33.76

5.6.4 Classification

We employed three pattern recognition algorithms, namely MLP, LDA and SVM for classification of 11 hand and wrist movements. MES data and classifiers description were given in chapter 4.6.2.

Table 5.5 compares the average accuracy between set 4 and a literature feature set across 20 subjects. In general, set 4 achieved higher accuracy than the other sets with respect to all classifiers. The MLP and LDA benefited from set 4 shown by the small p value. The SVM result for set 4, however, was not significantly

Table 5.4: Ranked features in final set according to CP index.

Ranking	Signal feature	RMSE ±SD	CP (%)
1	tren	0.079 ±0.023	26.27
2	tcr	0.074 ±0.028	17.11
3	tzc	0.069 ±0.028	7.49
4	tpks	0.067 ±0.027	7.33
5	tmpks & tmvel	0.065 ±0.027	4.72
6	tfdh	0.066 ±0.028	4.72
7	tslpch1	0.065 ±0.027	3.88
8	twl & tmfl	0.064 ±0.026	3.70
9	tfd	0.065 ±0.027	3.67
10	tstd	0.064 ±0.026	3.48
11	tdam	0.065 ±0.026	3.31
12	tmd	0.063 ±0.027	2.76

Table 5.5: Average accuracy of 11 hand movements of 20 subjects evaluated by three algorithms on set 4 and the literature feature set.

	Set 4 [a]		Literature set [b]		
Algorithm	Computational time (s)	acc (%) ±SD	Computational time (s)	acc (%) ±SD	Wilcoxon p-value
MLP	0.594 ±0.074	95.58 ±0.027	0.332 ±0.097	90.06 ±0.060	1.2042e-04
LDA	0.106 ±0.028	93.97 ±0.019	0.205 ±0.014	90.03 ±0.053	0.0012
SVM	0.152 ±0.015	95.95 ±0.016	0.233 ±0.015	95.42 ±0.024	0.1913

[a] *tzc, tcr, tren*
[b] *tmabs, twl, tzc, tslpch2*

different from the literature set. In term of computational time, set 4 took less time to compute the accuracy than the literature set when performed by LDA and SVM. When performed by MLP, set 4 took 0.594 seconds while the literature set took 0.332 seconds which was half of set 4.

In detail, Fig. 5.7 shows the accuracy measured on individual subject and individual movement. For MLP, set 4 was completely superior to the literature set both on subject and movement. For LDA, though there were some points that the literature outperformed set 4 on subject, the accuracy of set 4 on movement was completely higher than the literature set. There was an alternative leading between these sets performed by SVM across subjects. Set 4 was superior at some subjects such as subject 1, 7 while it was not at the other subjects for example subject 6, 9. It was very similar in overall accuracy on movement of these sets excluding some movements that set 4 was superior such as OH, PH, FG, and AG.

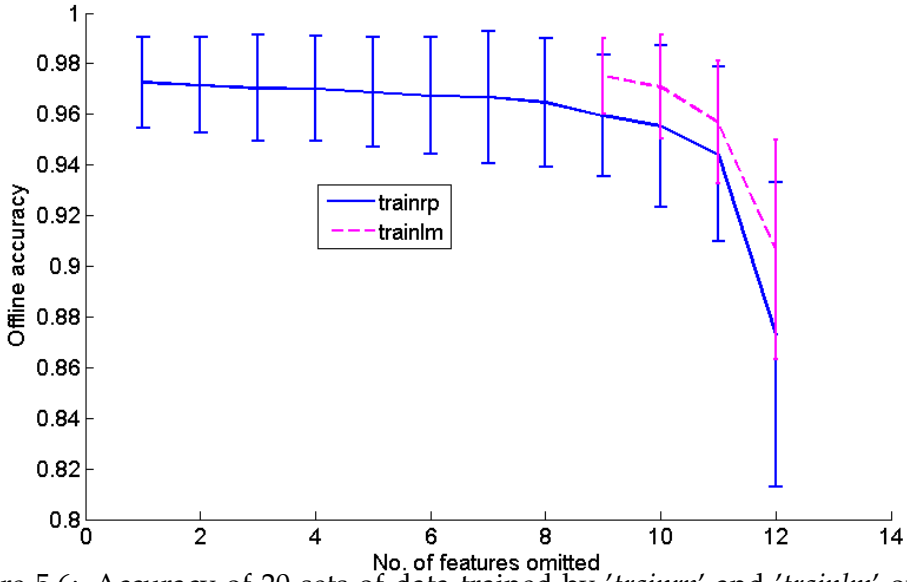


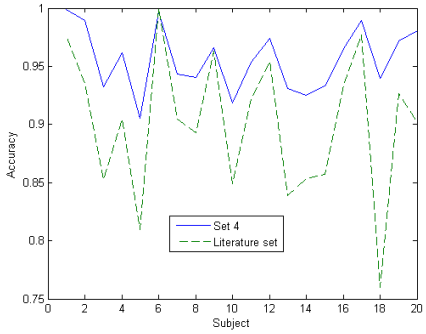
Figure 5.6: Accuracy of 20 sets of data trained by '*trainrp*' and '*trainlm*' over the omitted range of features.

5.7 Conclusion

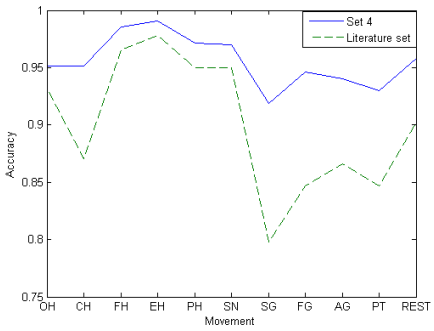
In this chapter, we present an FS approach using SOFM. We connect 122 time segments of potential signal features extracted from the MES waveform to time series vectors and use the SOFM to map these vectors to clusters. Dominant vectors in a cluster are selected and assembled with other vectors from other clusters. They are ranked by their CP values. The top ranking vectors are selected corresponding to the top signal features representing a feature set. This set is named as set 4 to differentiate with three feature sets previously drawn in chapter 4. The MLP, LDA and SVM demonstrated that set 4 was superior to a typical feature set found in the literature, particularly on MLP and LDA.

The SOFM is implemented by batch learning algorithms, which is more advantage than other methods in MATLAB R2013b. In addition, the network is optimised by initialisation of original neurons position and network dimension. The CP value is calculated after a signal feature being mapped by SOFM. This value together with a type of signal feature identified by SOFM are helpful to select an appropriate feature set.

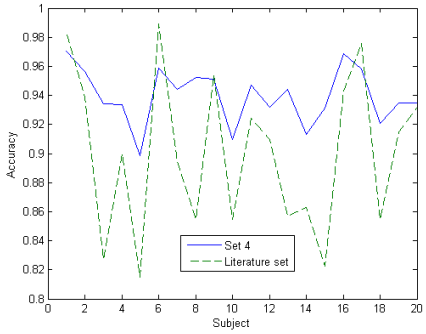
The main contribution of this chapter is that a time series analysis is first investigated providing a thorough examination on the behaviour of TD features influenced by muscle activity with respect to time. While the conventional FS approach compares signal features in segmented time windows, we are able to distinguish them based on their characteristics using time series analysis.



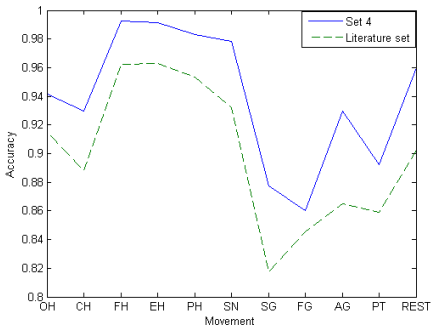
(a) MLP on subject



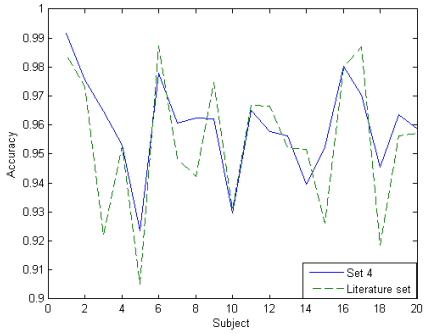
(b) MLP on movement



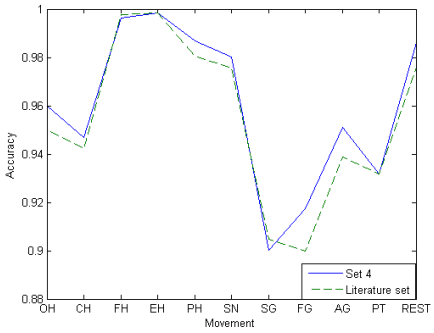
(c) LDA on subject



(d) LDA on movement



(e) SVM on subject



(f) SVM on movement

Figure 5.7: Accuracy of set 4 and literature set obtained by MLP, LDA and SVM measured on subject and movement.

Chapter 6

Conclusion

6.1 Thesis summary

This chapter highlights the main contributions of the research to the dimensionality reduction area of MES features in prosthetic control. The work has resulted alternative sets of features for pattern recognition in which feature selection is a key factor to implement the input. The typical TD features are considered in the thesis because of the ability to produce high classification rate of hand and wrist movements demonstrated in the literature. They are also simple to extract within a limit time frame which is suitable for real-time control. On one hand, there are several methods for dimensionality reduction, particularly by using feature projection method applying on TSD or TFD features. Those would likely result in a dramatic increment of the n-dimensional input vector. Thus, methodologies of projection have proved their benefits on TSD and TFD. When applying on TD features, those approaches have shown many disadvantages. On the other hand, feature selection have received few attention due to its poor performance when applied on TSD and TFD. Some typical TD features has been used directly to extract feature vectors for classification. There has been no standard study on the criteria of selected TD features. For this reason, the thesis presents alternatively methods of feature selection, providing logical argumentation in support of selecting TD features. The thesis is summarised as follows.

Chapter 1 provides an introduction about the research area of prosthetic control as well as motivation of the study. Motorised prostheses using the MES can be separated into two categories based on the control scheme namely the pattern recognition based and non-pattern recognition based. The former control system has demonstrated the ability to allow multi-function of the prosthetic arm, as opposed to the latter control scheme. Several pattern recognition algorithms have been developed over the past decades including MLPs, LDAs, SVMs, neural-fuzzy, hidden Markov models (HMMs), and Gaussian mixture models (GMMs). More sophisticated classifiers have been employed in the purpose of increasing the number of electrodes (up to 32) and hand/wrist movements (up to 10) as well as different types of data: steady-state, transient, continuous. Optimistic results report high classification rate (over 95%) on able body subjects and visual arms. On the other

hand, challenge remains in real-time control on the amputees due to congenital defects and muscle fatigue. The similar classification rate of the above classifiers both on able body subjects and visual arms suggests that those algorithms have reached a saturation point. This motivates us to focus on the implementation of the input which is the MES features. Typically, the MES features are categorised into four types: TD, TSD, FD, and TSCD or TFD. TD features are introduced at first due to the easy of simple mathematical expression and can be extracted by many types of microprocessors within a limit time frame. This guarantees the immediate response of the control scheme. Recent research has considered other types of features to enhance the quality of control. Several research has been taken in comparison of features. Although, many new features have recommended, TD features have consistently achieved the accuracy as high as others. In addition, the advantage of TD features is the extraction of lower input dimensionality. This becomes significantly important characteristics when prediction of required movement increases. Recently, feature projection methods have received considerable attention to reduce the input dimensionality of TSD and TFD features. TD features do not required input reduction since the input vector extracted by TD is lower dimensional vector. The limitation of TD features is one feature provides specific information of the muscle activity. It leads to a difficult task of classification of multifunctional movements. Thus, combination of many TD features which establishes a feature set is important. Since recommendation of sets of features has not been conducted in the literature, and there is limitation in the number of features in each study, these encourage us to focus on methods of finding optimal sets of features. This originally motivates the development of feature selection approach which preserves TD features providing valuable information and discard redundant features.

Chapter 2 substantiates the application of FS using MES features in the time domain by giving evidences in the literature. Signal features are extracted to feature vectors used as the input for the classifiers. A feature vector carries a portion of information in a segmented signal over the entire transient waveforms. The dimension of the vector depends on the number of features itself and the number of MES channels. When using TSD or TSC features, a feature produces a high dimensional vector required to transfer sufficient information. In addition, to adapt with the increment in the number of movements required in classification, researchers have utilised multiple MES channels together with various signal features. This results in an expression of enormous dimensional feature vectors. Recently, dimensionality reduction has received considerable attention in this area in an attempt to lower the input dimensionality which can be separated into two categories: FS and FP. Many studies on TSD or TSC features show that FP is a superior method. This leads to the concentration trend in FP. On the other hand, TD features which has demonstrated the ability to achieved classification rate as high as other types of features, and produces only one dimensional vector per channel, has received less attention. The input dimensionality of a feature vector individually depends on the number of MES channels. It significantly reduces the dimension of feature vectors while maintaining high accuracy. This supports the employment of FS which is suitable for TD features. Due to the fact that a minority of studies have been taken on FS, selectors have been proposed only by heuristic search engines in the literature. The representatives in this area are PSO, ACO and MI. These methods are applied to signal processing procedure in the same way as used for FP.

First, candidate features are presented to the network. Second, selector is employed to randomly determine sets of features. Third, feature vectors are extracted from those sets. Next, classification rate of feature sets are measured and the selector reserves the set corresponding to the highest classification rate as well as generates other random sets for competition. The limitations of this procedure are time consuming and local optimum trapped. The time limit for real-time control is 300 ms to avoid user frustration. Adding extra steps to signal processing would likely increase the time delay because the entire procedure also takes time to validate the network, leading to the disadvantages in real-time training. Local optimum is hard to avoid since the evaluation of the features relies on a particular classifier. Different classifiers vary the classification rate of the same dataset. To achieve the most benefit from a classifier, thorough validation process is required. However, it is occasionally taken due to time consuming. It can be seen that the attachment of selector affects the performance of the classifiers leading to the local optimum trapped. This raises the demand for development of a comprehensive selection approach. In the research aims, we introduce three FS approaches investigating other areas such as statistic criteria, unsupervised learning neural network and also the heuristic search. Apart from the heuristic search, the other methods are independent of classifiers allowing a fair evaluation of TD features.

Chapter 3 presents an alternative heuristic search using genetic algorithm (GA). Although this approach requires offline training to find an optimal set of features for individual subject, the GA is removed after the training stage allowing the control scheme to happen conventionally. We study twenty one typical TD features found in the literature. They are encoded twenty one genes in an artificial chromosome. A gene represents for the appearance of a particular feature, denoted by 1, or for the absence of that feature, denoted by 0. In other words, a chromosome is a binary string representing a feature set. The GA initialises a population of chromosomes, evaluates their performance and select the best fit chromosome for reproduction until the last generation reached. Crossover and mutation operators are responsible for exchanging genetic information and preserving diversity. For implementation, chromosome initialisation, and fitness evaluation are improved. The initial population of chromosomes is distributed evenly across all genetic zones to guarantee all features have equal possibility to be selected. The length of feature set also varies from one to twenty one features in a set, this ensures variety of combinations would be considered. Fitness evaluation employs fitness rank scaling function replacing the conventional raw scored fitness. It prevents dissemination of raw scored fitness by allowing evaluation takes place on the ranking of chromosomes in the population. It eliminates the probability of outperforming chromosomes overwhelm the others rapidly in case ordinary chromosomes may preserve beneficial information. Before practically applied to pattern recognition control system to find global maxima, the GA is studied under Rastrigins function to find global minimum, all parameters and operators remains. The Rastrigins function consists of several local minima but only one global minimum and the GA successfully demonstrates the ability to solve the problem. When running the GA on each subject, high classification rate is achieved approximately after twenty generations by a set of twelve features on average. The performance of these sets found by GA is superior a TD feature set found in the literature. It significantly improves the performance of low quality recorded data. The limitation is that GA acquires enormous features

to achieve the highest accuracy while the literature set achieves lower accuracy but only requires four features. The GA dramatically increases the number of features in a set opposing our purpose of finding concise sets. Consequently, we carry out statistical analysis to study the frequency of feature appearance in all subjects. Four features commonly appeared in selection of the GA are grouped to a set. However, this set performs poorly in the pattern recognition system. As a result, we conclude that heuristic search is not a solution for feature selection approach. It is either time consuming or input dimensional expanding.

Chapter 4 introduces a new method of feature selection using class relevant index, it measures the response of a feature to a target class of movements. We study three criteria namely mutual information, two-sample t-test and Mann-Whitney U test. These methods are widely applied in statistical test to analyse the probability distribution of data. It is inspired that the MES data also occurs in a certain probability density function. For this reason, we investigate MES data in three different ways. Mutual information (MI) examines the sharing knowledge between a feature and a target class, and selects a feature holding the largest contribution with that class, called Max-Relevance. In addition, we investigate the sharing knowledge between two features, and select a pair with the smallest contribution, called Min-Redundancy. A feature is selected by the highest subtraction of Max-Relevance and Min-Redundancy. The other criteria examine a feature by the ability to separate a target class to all the other classes. Two-sample t-test assumes that MES data are normally distributed while Mann-Whitney U test assumes the non-normal distributions of data. They are implemented by the complementary of correlation information and L2-norm distance between already selected features and candidate features. The advantages of those methods are short time examinations, and independence of pattern recognition algorithms. Three search engines result in three sets of features namely set 1, set 2, and set 3. We employ three classifiers which are the MLP, LDA and SVM to evaluate the performance of these sets. The same ratio of data is treated for all classifiers and optimal parameters are obtained in validation. Classification rate shows the superiority of the alternative sets as opposed to the set in literature. These sets can be used directly in training procedure and the method can be used for implementation of signal processing in real-time control.

Chapter 5 presents a non-linear clustering method using self-organising feature map (SOFM) to draw a feature set from twenty one potential TD features. Instead of a discrete time analysis as in chapter 4, we connect 122 segments of a signal feature to a transient waveform. Four input vectors are generated from this waveform, representing four MES channels of a signal feature. The SOFM analyses the input and maps it to a cluster. There are three types of cluster. Type 1 contains a completed set of four inputs in a cluster. This represents either identical signal features if more than two sets of input mapped to a cluster; or separated ones if there is only one set of input in a cluster. Type 2 comprises more than one input of a set in a cluster, representing shared information between signal features. Type 3 represents a signal feature that has only one input in a cluster while the other memberships are other signal features. This type of cluster together with the CP determines the importance of a signal feature in classification. A feature set, named as set 4, is drawn from twenty one potential TD feature. Set 4 consists of tzc, tcr, tren outperforming the literature feature set used in chapter 4. Classification performance is evaluated

by the MLP, LDA, and SVM which report consistently high accuracy on these classifiers. The statistical Wilcoxon p value indicates that the MLP and LDA benefited from set 4 while SVM did not. The advantage of non-linear SOFM is that the network is blind to target class. Therefore, a signal feature of this cluster is completely distinguished to that of other cluster making a high class separability of the input space. The improvement of the proposed method is also at evaluation of signal feature by the CP index which is measured by RMSEs. This provides a better evaluation since a large range of RMSEs can be targeted to a similar accuracy level. As a result, measurement by RMSEs is more accurate. In brief, set 4 together with other three feature sets proposed in chapter 4 can be alternative feature set in the time domain for hand and wrist classification. Moreover, the proposed method can be developed to experiment on real-time control schemes of motorised prostheses.

In conclusion, the project objectives addressed at the beginning of the thesis have been achieved. We review the development of dimensionality reduction methods to lower the input space of the MES data. Next, based on the limitations of recent research, we have developed three approaches investigating TD features in heuristic inspection, statistical analysis, and non-linear neural network. Then we have progressed three pattern recognition algorithms to evaluate the performance of feature sets produced by the proposed methods. We have demonstrated the ability of the alternative feature sets to achieve as high as or even higher classification rate as opposed to the literature feature set. Finally, we conclude that our proposed methods, in particular, the statistical analysis and non-linear neural network SOFM offers significant advantages in FS area with respect to TD features.

6.2 Thesis contributions

The major contribution of the thesis is to make a complete investigation of time domain features in order to develop methodologies for producing and evaluating feature sets. The proposed feature sets can be directly used in prosthetic control scheme without increasing the computational cost. The contributions compose of the following points.

6.2.1 Development of independent methods of feature selection

The study uses TD features which extract lower dimensional input vectors allowing dimensionality reduction to happen by preserving information rather than transforming them to a sub-input space. Therefore, selected features can be used directly for real-time control instead of repeating the procedure whenever the motorised prostheses execute. The advantage of TD features is simplicity and low computational cost which is suitable for a wide range of software and hardware application systems while maintaining high accuracy. The methods have the ability to accommodate with the increment of new features in the future by evaluating their performance to the already selected features.

6.2.2 Application of statistical analysis in feature selection

This is inspired by a simple but very interesting idea of investigating the relationship of data extracted from a particular feature to separate a target class. Generally, statistical analysis concerns all aspects of data in terms of constructing survey and experiments. However, this is the first time statistical analysis is applied in MES data to generate class relevant indices for signal features. Implemented by correlation and class separability coefficient, statistical criteria provide a powerful measurement of signal features to the classification rate. It saves the time consumed in order to either directly collaborate to the control scheme or being a stand-alone application in the offline training stage.

6.2.3 Investigation of MES data along transient waveform

The conventional methods of analysing MES data are to segment the entire signal waveform into segmentations of data. Each segment is considered as an observation or an input sample of the network and signal features are computed in each segment. In the last approach, we conduct a study of integrated signal features, joining separated time segments of a signal feature to a time series vector. This allows us to analyse the behaviour of signal features across the continuous stage of muscle activity, making it easy for features clustering.

6.3 Suggestion for future work

Although there is a high potential in application of the alternative feature sets as well as application of the proposed methods in pattern recognition based prosthetic control, some additional ideas may be worth sharing in regard to future research and development.

6.3.1 Application in real-time control

The limitation of the experimental method is data was provided by other research group, and is subject to their bias on collection and interpretation of data. Development of a data acquisition system and real-time experimentation will broaden data sources to construct an exhaustive database and compose real-time evaluation by visual and real motorised prostheses. Classification rates reported by most of recent studies is analysed from data of able body subjects, this reduces the benefit pointing to the amputees in purpose. However, due to the limitations of current technology, treatment and experimentation on amputees have remained challenging. By conducting a simple study approach, we hope to make contributions to the research community in the diversity of real-time data resources as well as ideas to take a closer step to a real-time control by amputees.

6.3.2 Clustering datasets

During experiments, we realised that there is no global optimal feature set suitable for all subjects and it is challenging to find a particular set for an individual subject. Consequently, datasets acquired from all volunteers can be categorised into clusters. It is convenient and feasible to produce a particular feature set for a cluster of data. Data clustering can be constructed by characteristics of MES features such as absolute values of signal amplitudes, frequency or transformation of the transient waveform. Subsequently, conventional analysis process can be applied on datasets of a cluster to search for an optimal feature set. This prevents the bias of averaging performance that typically ignores the outlier datasets.

BIBLIOGRAPHY

[1] AH. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram. A preliminary investigation of the effect of force variation for myoelectric control of hand prosthesis. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 5758–5761, July 2013.

[2] Hanan S Alghamdi, H Lilian Tang, and Saleh Alshomrani. Hybrid aco and tofa feature selection approach for text classification. In *Evolutionary Computation (CEC), 2012 IEEE Congress on*, pages 1–6. IEEE, 2012.

[3] Sridhar Arjunan and Dinesh Kumar. Decoding subtle forearm flexions using fractal features of surface electromyogram from single and multiple sensors. *Journal of neuroengineering and rehabilitation*, 7(1):53, 2010.

[4] Mohammadreza Asghari Oskoei and Huosheng Hu. Myoelectric control systemsa survey. *Biomedical Signal Processing and Control*, 2(4):275–294, 2007.

[5] Richard Bellman, Richard Ernest Bellman, Richard Ernest Bellman, and Richard Ernest Bellman. *Adaptive control processes: a guided tour*, volume 4. Princeton University Press Princeton, 1961.

[6] Asa Ben-Hur, David Horn, Hava T. Siegelmann, and Vladimir Vapnik. Support vector clustering. *J. Mach. Learn. Res.*, 2:125–137, March 2002.

[7] Christopher M Bishop et al. Neural networks for pattern recognition. 1995.

[8] Reza Boostani and Mohammad Hassan Moradi. Evaluation of the forearm emg signal features for the control of a prosthetic hand. *Physiological measurement*, 24(2):309, 2003.

[9] Klaus Buchenrieder. Processing of myoelectric signals by feature selection and dimensionality reduction for the control of powered upper-limb prostheses. In *Computer Aided Systems Theory–EUROCAST 2007*, pages 1057–1065. Springer, 2007.

[10] CHRISTOPHER J.C. BURGES. A tutorial on support vector machines for pattern recognition.

[11] Jörg Butterfaß, Markus Grebenstein, Hong Liu, and Gerd Hirzinger. Dlr-hand ii: Next generation of a dextrous robot hand. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 1, pages 109–114. IEEE, 2001.

[12] R. Chattopadhyay, G. Pradhan, and S. Panchanathan. Subject independent computational framework for myoelectric signals. In *Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE*, pages 1–4, May 2011.

[13] Jun-Uk Chu, Inhyuk Moon, and Mu seong Mun. A real-time emg pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand. *Biomedical Engineering, IEEE Transactions on*, 53(11):2232–2239, Nov 2006.

[14] Jun-Uk Chu, Inhyuk Moon, and Mu seong Mun. A supervised feature projection for real-time multifunction myoelectric hand control. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 2417–2420, Aug 2006.

[15] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[16] David L. Davies and Donald W. Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-1(2):224–227, April 1979.

[17] Lawrence David Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York, 1991.

[18] Janez Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1 – 30, 2006.

[19] J. A. Dutton. *The Ceaseless Wind: An Introduction to the Theory of Atmospheric Motion*. 1976.

[20] Verona Butu Suzana Cismas Mihai Lungu Paul Schiopu Adrian Barbilian Anca Plavitu Eduard Franty, Lucian Milea. Methods of acquisition and signal processing for myoelectric control of artificial arms. *Romanian journal of information science and technology*, 2012.

[21] K. Englehart, B. Hudgin, and P.A. Parker. A wavelet-based continuous classification scheme for multifunction myoelectric control. *Biomedical Engineering, IEEE Transactions on*, 48(3):302–311, March 2001.

[22] K. Englehart, B. Hudgins, P. Parker, and M. Stevenson. Time-frequency representation for classification of the transient myoelectric signal. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 5, pages 2627–2630 vol.5, Oct 1998.

[23] Kevin Englehart, B Hudgin, and Philip A Parker. A wavelet-based continuous classification scheme for multifunction myoelectric control. *Biomedical Engineering, IEEE Transactions on*, 48(3):302–311, 2001.

[24] Kevin Englehart, Bernard Hudgins, Philip A Parker, and Maryhelen Stevenson. Classification of the myoelectric signal using time-frequency based representations. *Medical engineering & physics*, 21(6):431–438, 1999.

[25] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt. A comparison of waveform fractal dimension algorithms. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 48(2):177–183, 2001.

[26] Kristin A Farry, Ian D Walker, and Richard G Baraniuk. Myoelectric teleoperation of a complex robotic hand. *Robotics and Automation, IEEE Transactions on*, 12(5):775–788, 1996.

[27] Daniel Graupe and William K Cline. Functional separation of emg signals via arma identification methods for prosthesis control purposes. *Systems, Man and Cybernetics, IEEE Transactions on*, (2):252–259, 1975.

[28] Daniel Graupe, Javad Salahi, and Kate H Kohn. Multifunctional prosthesis and orthosis control via microcomputer identification of temporal pattern differences in single-site myoelectric signals. *Journal of Biomedical Engineering*, 4(1):17–22, 1982.

[29] T Higuchi. Approach to irregular time series on the basis of the fractal theory. *Physica D*, 31:277 – 283, 1988.

[30] N. Hogan and Robert W. Mann. Myoelectric signal processing: Optimal estimation applied to electromyography - part i: Derivation of the optimal myoprocessor. *Biomedical Engineering, IEEE Transactions on*, BME-27(7):382–395, July 1980.

[31] N. Hogan and Robert W. Mann. Myoelectric signal processing: Optimal estimation applied to electromyography - part ii: Experimental demonstration of optimal myoprocessor performance. *Biomedical Engineering, IEEE Transactions on*, BME-27(7):396–410, July 1980.

[32] Han-Pang Huang, Yi-Hung Liu, Li-Wei Liu, and Chin-Shin Wong. Emg classification for prehensile postures using cascaded architecture of neural networks with self-organizing maps. In *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, volume 1, pages 1497–1502 vol.1, Sept 2003.

[33] Hu Huang, Hong-Bo Xie, Jing-Yi Guo, and Hui-Juan Chen. Ant colony optimization-based feature selection method for surface electromyography signals classification. *Computers in biology and medicine*, 42(1):30–38, 2012.

[34] Bernard Hudgins, Philip Parker, and Scott Robert N. A new strategy for multifunction myoelectric control. *IEEE transactions on bio-medical engineering*, 40(1):82–94, 1993.

[35] GIDEON F Inbar and Antonie E Noujaim. On surface emg spectral characterization and its application to diagnostic classification. *Biomedical Engineering, IEEE Transactions on*, (9):597–604, 1984.

[36] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1):4–37, 2000.

[37] Richard Jensen. *Combining rough and fuzzy sets for feature selection*. PhD thesis, Citeseer, 2005.

[38] Michael J. Katz. Fractals and the analysis of waveforms. *Computers in Biology and Medicine*, 18(3):145–156, 1988.

[39] L.M.D. Khong, T.J. Gale, D. Jiang, J.C. Olivier, and M. Ortiz-Catalan. Multi-layer perceptron training algorithms for pattern recognition of myoelectric signals. In *Biomedical Engineering International Conference (BMEiCON), 2013 6th*, pages 1–5, Oct 2013.

[40] Rami N Khushaba and Adel Al-Jumaily. Channel and feature selection in multifunction myoelectric control. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 5182–5185. IEEE, 2007.

[41] Rami N Khushaba, Akram AlSukker, Ahmed Al-Ani, Adel Al-Jumaily, and Albert Y Zomaya. A novel swarm based feature selection algorithm in multifunction myoelectric control. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 20(4, 5):175–185, 2009.

[42] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(13):1 – 6, 1998.

[43] Martin A Kraaijveld, Jianchang Mao, and Anil K Jain. A nonlinear projection method based on kohonen's topology preserving maps. *Neural Networks, IEEE Transactions on*, 6(3):548–559, 1995.

[44] Todd A Kuiken, Guanglin Li, Blair A Lock, Robert D Lipschutz, Laura A Miller, Kathy A Stubblefield, and Kevin B Englehart. Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms. *Jama*, 301(6):619–628, 2009.

[45] Peter J Kyberd, Owen E Holland, Paul H Chappell, Simon Smith, Robert Tregidgo, Paul J Bagwell, and Martin Snaith. Marcus: A two degree of freedom hand prosthesis with hierarchical grip control. *Rehabilitation Engineering, IEEE Transactions on*, 3(1):70–76, 1995.

[46] Guanglin Li, Aimee E Schultz, and Todd A Kuiken. Quantifying pattern recognition-based myoelectric control of multifunctional transradial prostheses. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(2):185–192, 2010.

[47] Xiangjun Li and Fen Rao. An rough entropy based approach to outlier detection. *Journal of Computational Information Systems*, 8(24):10501–10508, 2012.

[48] CM Light, PH Chappell, B Hudgins, and K Engelhart. Intelligent multifunction myoelectric control of hand prostheses. *Journal of Medical Engineering & Technology*, 26(4):139–146, 2002.

[49] D. Lowe and A.R. Webb. Optimized feature extraction and the bayes decision in feed-forward classifier networks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(4):355–364, Apr 1991.

[50] Howard B. Demuth Mark Hudson Beale, Martin T. Hagan. *Neural Network Toolbox, Users Guide, R2013b*. MathWorks.

[51] D. Randall Wilson; Tony R. Martinez. The general inefficiency of batch training for gradient descent learning. *Neural Networks 16*, 2003.

[52] Roberto Merletti, L. Lo Conte, E. Avignone, and P. Guglielminotti. Modeling of surface myoelectric signals. i. model implementation. *Biomedical Engineering, IEEE Transactions on*, 46(7):810–820, July 1999.

[53] Silvestro Micera, Jacopo Carpaneto, and Stanisa Raspopovic. Control of hand prostheses using peripheral information. *IEEE Reviews in Biomedical Engineering*, 3(1):48–68, 2010.

[54] Silvestro Micera, Angelo M Sabatini, and Paolo Dario. On automatic identification of upper-limb movements using small-sized training sets of emg signals. *Medical engineering & physics*, 22(8):527–533, 2000.

[55] Melanie Mitchell. *An introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996.

[56] Ganesh Naik and Dinesh Kumar. Hybrid feature selection for myoelectric signal classification using mica. *Journal of Electrical Engineering*, 61(2):93–99, 2010.

[57] Xavier Navarro, Thilo B. Krueger, Natalia Lago, Silvestro Micera, Thomas Stieglitz, and Paolo Dario. A critical review of interfaces with the peripheral nervous system for the control of neuroprostheses and hybrid bionic systems. *Journal of the Peripheral Nervous System*, 10(3):229–258, 2005.

[58] Max Ortiz-Catalan. Biopatrec, 2013.

[59] Max Ortiz-Catalan, Rickard Branemark, and Bo Hakansson. Biopatrec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms. *Source Code for Biology and Medicine*, 8(1):11, 2013.

[60] Mohammadreza Asghari Oskoei and Huosheng Hu. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE transactions on biomedical engineering*, 55(8):19561965, 2008.

[61] Chih-Jen Lin Pai-Hsuen Chen and Bernhard Scholkopf. A tutorial on nu-support vector machines.

[62] Sankar K Pal, B Uma Shankar, and Pabitra Mitra. Granular computing, rough entropy and object extraction. *Pattern Recognition Letters*, 26(16):2509–2517, 2005.

[63] Sang-Hui Park and Seok-Pil Lee. Emg pattern recognition based on artificial intelligence techniques. *Rehabilitation Engineering, IEEE Transactions on*, 6(4):400–405, 1998.

[64] H. Peng, Fulmi Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, Aug 2005.

[65] Angkoon Phinyomark, Chusak Limsakul, and Pornchai Phukpattaranont. A novel feature extraction for robust emg pattern recognition. *Journal of computing*, 2009.

[66] Naoki Saito and Ronald R Coifman. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5(4):337–358, 1995.

[67] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[68] Ann M. Simon, Levi J. Hargrove, Blair A. Lock, and Todd A. Kuiken. Target achievement control test: Evaluating realtime myoelectric pattern-recognition control of multifunctional upper-limb prostheses. *Journal of Rehabilitation Research & Development*, 48(6):619–628, 2011.

[69] M. Srinivas and L.M. Patnaik. Adaptive probabilities of crossover and mutation in genetic algorithms. *Systems, Man and Cybernetics, IEEE Transactions on*, 24(4):656–667, Apr 1994.

[70] Francesco VG Tenore, Ander Ramos, Amir Fahmy, Soumyadipta Acharya, Ralph Etienne-Cummings, and Nitish V Thakor. Decoding of individuated finger movements using surface electromyography. *Biomedical Engineering, IEEE Transactions on*, 56(5):1427–1434, 2009.

[71] J. Thuburn and Y. Li. Numerical simulations of rossby-haurwitz waves. 52:181–189, 2000.

[72] Dennis Tkach, He Huang, and Todd A Kuiken. Research study of stability of time-domain features for electromyographic pattern recognition. *J Neuroeng Rehabil*, 7:21, 2010.

[73] A. Ultsch and H. P. Siemon. Kohonens self organizing feature maps for exploratory data analysis. *Neural Network Conf*, pages 305–308, 1990.

[74] Jason Weston. Support vector machine (and statistical learning theory) tutorial.

[75] D. L. Williamson, J. B. Drake, J. J. Hack, R. Jakob, and Swarztrauber P. N. A standard test set for numerical approximations to the shallow water equations in spherical geometry. 102:211–224, 1992.

[76] Stephen Winters-Hilt and Sam Merat. Svm clustering. *BMC bioinformatics*, 8(Suppl 7):S18, 2007.

[77] Mahyar Zardoshti-Kermani, Bruce C Wheeler, Kambiz Badie, and Reza M Hashemi. Emg feature evaluation for movement control of upper extremity prostheses. *Rehabilitation Engineering, IEEE Transactions on*, 3(4):324–333, 1995.

[78] Micera Zecca, Silvestro Micera, MC Carrozza, and P Dario. Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews in Biomedical Engineering*, 30(4-6), 2002.

[79] J Zhong, J Shi, Y Cai, and Q Zhang. Recognition of hand motions via surface emg signal with rough entropy. *33rd Annu. Int. Conf. IEEE EMBS*, pages 4100 – 4103, 2011. 30 Aug - 3 Sep.